Nonconvex optimization for improved exploitation of gradient sparsity in CT image reconstruction

Sidky, Emil Y.; Chartrand, Rick; Jørgensen, Jakob Sauer; Pan, Xiaochuan

Published in:
Proceedings of 12th Fully 3D Meeting - 2013

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Tahoe Rim Trail, ©2012 cariberry@flickr.com

Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine

Proceedings

June 16-21, 2013
Granlibakken Resort, Lake Tahoe, California

Welcome Message

Welcome to the 2013 International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine which is being held at the Granlibakken Conference Center & Lodge near Lake Tahoe in California.

This is the 12th in a series of meetings that have served as one of the major forums for presentation of new results in the field of 3D image reconstruction, primarily with applications in x-ray computer tomography, PET and SPECT. The proceedings of the 2013 and all past meetings are archived at http://www.fully3d.org.

Over the life-time of the meeting the focus has shifted to reflect recent developments in the field. Many of the major developments in fully 3D PET and SPECT imaging were first presented at Fully3D, as were the key results for analytic reconstruction methods in cone beam x-ray CT. While a broad range of topics are represented at the current meeting, method for low-dose x-ray CT are clearly a major focus of many of the papers at the current meeting. Recent meetings have included a separate one day workshop on High Performance Computing for Fully3D imaging. For 2013 we decided to integrate this topic as a regular component of the meeting, and you will find an oral session as well as many poster papers on this important topic. On Monday night after dinner we invite you to attend a panel discussion led by experts from the major imaging companies on the current status and future of 3D imaging.

Fully3D has always been an independent meeting and we have continued this tradition. We are therefore particularly grateful to our sponsors (listed on the next page) for their valuable financial support. We would also like to express our appreciation to the Scientific Committee for their prompt reviews of the large number of papers submitted to the meeting and to the members of the Organizing Committee for their invaluable help. Finally, our thanks to Gloria Halfacre and Seth Scafani (USC), Susie Helton (UC Davis), and the staff at the Granlibakken who have all been of great assistance in organizing this meeting.

Social events during the week include a reception on Sunday evening, a dinner cruise on Lake Tahoe on Tuesday and a barbeque on Wednesday evening. We hope you enjoy the technical program and the opportunity to socialize with friends and colleagues here in the beautiful Sierra Nevada mountains.

Richard Leahy and Jinyi Qi
Conference Chairs

Sponsors

PHILIPS

SIEMENS

TOSHIBA

Leading Innovation

GE Healthcare

Ming Hsieh Institute

Ming Hsieh Department of Electrical Engineering

USC Viterbi

School of Engineering

UC DAVIS

COLLEGE OF ENGINEERING

Conference Organization

Conference Chairs

Richard M. Leahy	University of Southern California
Jinyi Qi	University of California Davis

Organizing Committee

Evren Asma	GE Global Research, NY
Bing Bai	University of Southern California
Ron Huesman	Lawrence Berkeley Lab
Quanzheng Li	Massachusetts General Hospital
Guobao Wang	University of California Davis
Jian Zhou	University of California Davis

Administrative and Support Staff

Gloria Halfacre	University of Southern California
Susie Helton	University of California Davis

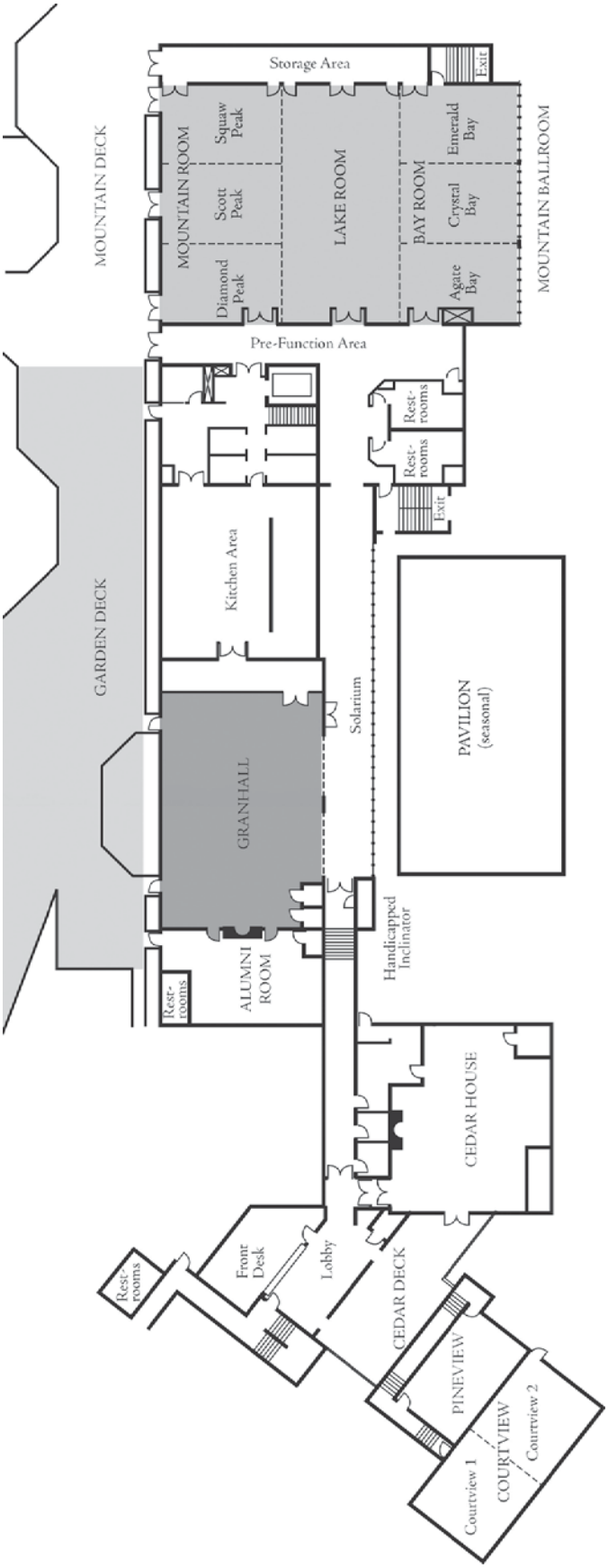
Student Volunteers

Kuang Gong	University of California Davis
Yanguang Lin	University of Southern California
Li Yang	University of California Davis
Mengxi Zhang	University of California Davis
Wentao Zhu	University of Southern California

Scientific Committee

Alexander Zamyatin	Toshiba Medical Systems, USA
Andrew Reader	Montreal Neurological Institute, Canada
Bing Bai	University of Southern Calfornia, USA
Benjamin M. W. Tsui	Johns Hopkins Hospital, USA
Bruno DeMan	GE Global Research, USA
Carl R. Crawford	Csuptwo, LLC, USA
Dan J. Kadrmas	University of Utah, USA
David Kaeli	North Eastern University, USA
Evren Asma	GE Global Research, USA
Franz Pfeiffer	TU München, Germany
Frédéric Noo	University of Utah, USA
Freek Beekman	University Hospital Utrecht, Netherlands
Ge Wang	Virginia Tech, USA
Grant T. Gullberg	Lawrence Berkeley National Laboratory,USA
Guang-Hong Chen	University of Wisconsin-Madison,USA
Guobao Wang	UC Davis, USA
Günter Lauritsch	Siemens Healthcare, Germany
Jean-Baptiste Thibault	GE Healthcare Technologies, WI
Hiroyuki Kudo	University of Tsukuba, Japan
Jason Corso	University at Buffalo SUNY, USA
Jeffrey A. Fessler	University of Michigan, USA
Jens Gregor	The University of Tennessee, USA
Jiang Hsieh	GE Healthcare, USA
Johan Nuyts	Katholieke Universiteit Leuven, Belgium
Ken D. Sauer	University of Notre Dame, USA
Katsuyuki Taguchi	The Johns Hopkins Hospital, USA
Klaus Mueller	Stony Brook University, USA
Larry Zeng	University of Utah, USA
Laurent Desbat	Université Joseph Fourier, France
Magdalena Rafecas	University of Valencia, Spain
Marc Kachelrieß	German Cancer Research Center (DKFZ), Germany
Michael King	The University of Massachusetts, USA
Michael Silver	Toshiba Medical Systems, USA
Michel Defrise	Vrije Universiteit Brussels, Belgium
Nicole Maass	Siemens Healthcare, Germany
Patrick La Riviere	University of Chicago, USA
Paul E. Kinahan	University of Washington, USA
Per-Erik Danielsson	Linköping University, Sweden
Roland Proksa	Philips Medical Systems, Germany
Ronald Huesman	Lawrence Berkeley National Laboratory,USA
Sarah Patch	University of Wisconsin-Milwaukee, USA
Stefaan Vandenberghe	Ghent University, Belgium
Thomas Flohr	Siemens Healthcare, Germany
Thomas Koehler	Philips Research Europe, Germany
Xiangyang Tang	Emory University, USA
Xiaochuan Pan	University of Chicago,USA
Yuxiang Xing	Tsinghua University, Beijing, China
Zhengrong Ying	Zomographic, USA
Wenli Wang	Toshiba Medical Systems, USA

Resort Map


Storage Area
Exit
MOUNTAIN DECK
MOUNTAIN ROOM
Squaw Peak
Scott Peak
Diamond Peak
LAKE ROOM
BAY ROOM
Emerald Bay
Crystal Bay
Agate Bay
MOUNTAIN BALLROOM
Pre-Function Area
Rest-rooms
Rest-rooms
Exit
Kitchen Area
GARDEN DECK
Solarium
PAVILION (seasonal)
GRANHALL
Handicapped Inclinator
ALUMNI ROOM
Rest-rooms
CEDAR HOUSE
Rest-rooms
Front Desk
Lobby
CEDAR DECK
PINEVIEW
COURTVIEW
Courtview 2
Courtview 1

Sunday, June 16

17.00 - 19.00	Registration	
18.00 - 21.00	Welcome Reception	Garden Deck

Monday, June 17

07.00 - 08.00	Registration/Breakfast	Granhall
08.00 - 08.20	Opening Remarks	Mountain/Lake
08.20 - 10.00	MoO1 CT Reconstruction I Session Chair: Xiaochuan Pan	Mountain/Lake
10.00 - 10.30	Coffee	Pre-Function Area
10.30 - 11.50	MoO2 Motion Estimation Session Chair: Grant T. Gullberg	Mountain/Lake
12.00 - 13.30	Lunch	Garden Deck
13.30 - 15.30	MoP1 Posters Session Chair: Johan Nuyts	Mountain/Lake
15.30 - 16.00	Coffee	Pre-Function Area
16.00 - 17.40	MoO3 PET Reconstruction Session Chair: Paul E. Kinahan	Mountain/Lake
18.00 - 19.00	Cocktails	Garden Deck
19.00 - 20.00	Industry Panel Discussion	Mountain/Lake
20.00 - 22.00	Dinner	Garden Deck

In event of inclement weather, all Garden Deck activities will be relocated to Granhall

Tuesday, June 18

07.15 - 08.20	Breakfast	Granhall
08.20 - 10.00	TuO1 CT Reconstruction II Session Chair: Yuxiang Xing	Mountain/Lake
10.00 - 10.30	Coffee	Pre-Function Area
10.30 - 11.50	TuO2 Dynamic Methods Session Chair: Charles A. Bouman	Mountain/Lake
12.00 - 13.30	Lunch	Granhall
13.30 - 16.30	Afternoon Activities/Free Time	
16.30	Bus to Dinner	Lobby
18.00 - 21.30	Lake Tahoe Dinner Cruise	*Tahoe Queen*
21.30	Bus to Granlibakken Resort	Pier

In event of inclement weather, all Garden Deck activities will be relocated to Granhall

Wednesday, June 19

07.15 - 08.20	Breakfast	Granhall
08.20 - 10.00	WeO1 High Performance Computing Session Chair: Marc Kachelrieß	Mountain/Lake
10.00 - 10.30	Coffee	Pre-Function Area
10.30 - 11.50	WeO2 Acquisition Geometries Session Chair: Frédéric Noo	Mountain/Lake
12.00 - 13.30	Lunch	Garden Deck
13.30 - 15.30	WeP2 Posters Session Chair: Benjamin M. W. Tsui	Mountain/Lake
15.30 - 16.00	Coffee	Pre-Function Area
16.00 - 17.40	WeO3 System Design and Calibration Session Chair: Michael A. King	Mountain/Lake
18.00 - 19.00	Cocktails	Garden Deck
19.00 - 21.00	BBQ	Garden Deck

In event of inclement weather, all Garden Deck activities will be relocated to Granhall

Thursday, June 20

07.15 - 08.20	Breakfast	Granhall
08.20 - 10.00	ThO1 PET & SPECT Session Chair: Michel Defrise	Mountain/Lake
10.00 - 10.30	Coffee	Pre-Function Area
10.30 - 11.50	ThO2 Artifact Reduction and Postprocessing Session Chair: Jiang Hsieh	Mountain/Lake
12.00 - 13.30	Lunch	Garden Deck
13.30 - 15.30	ThP3 Posters Session Chair: Wenli Wang	Mountain/Lake
15.30 - 16.00	Coffee	Pre-Function Area
16.00 - 17.40	ThO3 Regularization Methods Session Chair: Jeffrey A. Fessler	Mountain/Lake
18.00 - 19.30	Cocktails	Garden Deck
19.30 - 21.00	Dinner	Garden Deck

In event of inclement weather, all Garden Deck activities will be relocated to Granhall

Friday, June 21

07.30 - 09.00	Breakfast	Granhall
10.30	Granlibakken Resort Checkout	Lobby

Monday, June 17

Oral Session: CT Reconstruction I (8:20 - 10:00 AM)

Session Chair: Xiaochuan Pan

8:20 - 8:40	*Donghwan Kim, Sathish Ramani, and Jeffrey A. Fessler*: Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods……p22.
8:40 - 9:00	*Ludwig Ritschl, Michael Knaup and Marc Kachelrieß*: Extending the Dynamic Range of Flat Detectors in CBCT using a Compressed - Sensing–Based Multi–ExposureTechnique……p26.
9:00 - 9:20	*Il Yong Chun and Thomas M. Talavage* Efficient Compressed Sensing Statistical X-Ray/CT Reconstruction from Fewer Measurements……p30.
9:20 - 9:40	*Jingyan Xu and Benjamin M. W. Tsui* C-Arm CT Image Reconstruction From Sparse Projections……p34.
9:40 - 10:00	*Helene Langet, Aymeric Reshef, Cyril Riddell, Yves Trousset, Arthur Tenenhaus, Elisabeth Lahalle, Gilles Fleury, and Nikos Paragios*: Nonlinear diffusion constraints for reconstructing subsampled rotational angiography data……p38.

Oral Session: Motion Estimation (10:30 - 11:50 AM)

Session Chair: Grant T. Gullberg

10:30 -10:50	*Marcus Brehm, Pascal Paysan, Markus Oehlhafen, and Marc Kachelrieß* : Robust Motion Estimation for On–Board CBCT Imaging using an Angular Sampling Artifact Model……p42.
10:50 - 11:10	*Qiulin Tang, Jochen Cammin, and Katsuyuki Taguchi*: Four-dimensional projection-based motion estimation and compensation for cardiac x-ray computed tomography……p46.
11:10 - 11:30	*Chris Schwemmer, Christopher Rohkohl, Günter Lauritsch*, *Kerstin Müller, and Joachim Hornegger*: Opening Windows- Increasing Window Size in Motion-Compensated ECG-gated Cardiac Vasculature Reconstruction ……p50.
11:30 - 11:50	*Wolfgang Wein and Alexander Ladikos*: Towards General Motion Recovery in Cone-Beam Computed Tomography……p54.

Oral Session: PET Reconstruction (16:00 - 17:40 PM)

Session Chair: Paul E. Kinahan

16:00 - 16:20	*Vladimir Y. Panin, Michel Defrise and Michael E. Casey*: TOF ML-ACF reconstruction using CT based attenuation as a priori information…..…..p58.
16:20 - 16:40	*Se Young Chun, Yuni K. Dewaraja, and Jeffrey A. Fessler*: Alternating Direction Method of Multiplier for Emission Tomography with Non Local Regularizers…….…p62.
16:40 - 17:00	*Michel Defrise, Ahmadreza Rezaei, Johan Nuyts*: Simultaneous reconstruction of attenuation and activity in TOF-PET: analysis of the convergence of the MLACF algorithm……..……..….p66.
17:00 - 17:20	*Guobao Wang and Jinyi Qi*: Edge-Preserving PET image reconstruction using trust optimization transfer……..…..p70.
17:20 - 17:40	*Thomas Kosters, Michael Fieseler, Tobias Block, Frank Wubbeling, David Faul, Fernando Boada, Klaus Schafers*: Combined AW-OSEM Reconstruction and Mass-Preserving Motion Correction of PET Data……p74.

Poster Session (13:30 - 15:30 PM)

Session Chair: Johan Nuyts

- *Jang Hwan Cho and Jeffrey A. Fessler*: Quadratic regularization design for 3D axial CT……..……..……..…….……....p78.
- *Steven Oeckl*: CT Reconstruction of 3D Wavelet Coefficients and its Application to Nondestructive Testing……..……..…….…….....p82.
- *Junguo Bian, Kai Yang, Xiao Han, Emil Y. Sidky, John M. Boone, and Xiaochuan Pan*: Constrained TV-minimization Reconstruction in Low-dose Breast CT ……..…..p86.
- *Jian Zhou, Katherine L. Walker, Gregory S. Mitchell, Simon R. Cherry and Jinyi Qi*: Maximum A Posteriori Image Reconstruction for A High Sensitivity Uncollimated Small-Animal SPECT System……..…....p90.
- *Abhinav K. Jha, Harrison H. Barrett, Eric Clarkson, Luca Caucci, and Matthew A.* Kupinski: Analytic methods for list-mode reconstruction ……..……p94.
- *Kerstin Müller, Chris Schwemmer, Günter Lauritsch, Christopher Rohkohl, Andreas K. Maier, Hein Heidbüchel, Stijn De Buck, Dieter Nuyens, Yiannis Kyriakou, Christoph Köhler, Rebecca Fahrig, and Joachim Hornegger:* Image Artifact Influence on Motion Compensated Tomographic Reconstruction in Cardiac C-arm CT ..……p98.
- *Yunlong Zan, Rostyslav Boutchko, Qiu Huang, and Grant T. Gullberg*: 4-D Reconstruction of SPECT 123I-MIBG Data Acquired with Slow-Rotation Cameras via Spatial-Temporal

Tuesday, June 18

Oral Session: CT Reconstruction II (8:20 - 10:00 AM)

Session Chair: Yuxiang Xing

8:20 - 8:40	*Emil Y. Sidky, Rick Chartrand, Jakob S. Jørgensen, and Xiaochuan Pan*: Nonconvex optimization for improved exploitation of gradient sparsity in CT image reconstruction.....p189.
8:40 - 9:00	*Barbara Flach, Jan Kuntz, Marcus Brehm, Rolf Kueres, Sonke Bartling, and Marc Kachelrieß*: Up–To–Date Prior Knowledge via Motion Correction for Low Dose Tomographic Fluoroscopy.....p193.
9:00 - 9:20	*Zhou Yu, Lin Fu, Debashish Pal, Jean-Baptiste Thibault, Charles A. Bouman and Ken D. Sauer*: Nested Loop Algorithm for Parallel Model Based Iterative Reconstruction.....p197.
9:20 - 9:40	*Adam S. Wang, J. Webster Stayman, Yoshito Otake, Gerhard Kleinszig, Sebastian Vogt, A. Jay Khanna, Ziya L. Gokaslan, Jeffrey H. Siewerdsen*: Statistical Reconstruction for Soft Tissue Imaging with Low Dose C-arm Cone-Beam CT.....p201.
9:40 - 10:00	*Lin Fu, Zhou Yu, Jean-Baptiste Thibault, Bruno De Man, Madison G. McGaffin, and Jeffrey A. Fessler*: Space-Variant Channelized Preconditioner Design for 3D Iterative CT Reconstruction.....p205.

Oral Session: Dynamic Methods (10:30 - 11:50 AM)

Session Chair: Charles A. Bouman

10:30 -10:50	*Uttam Shrestha, Fares Alhassen, Rostyslav Boutchko, Robert G. Gould, Youngho Seo, Elias H. Botvinick, Grant T. Gullberg*: Fully 6D image reconstruction for myocardial perfusion imaging of tracer dynamics, cardiac and respiratory motion.....p209.
10:50 - 11:10	*Wentao Zhu, Bing Bai, Peter S. Conti, Quanzheng Li, Richard M. Leahy*: Data Correction Methods for Wholebody Patlak Imaging from List-mode PET Data.p213.
11:10 - 11:30	*Dimple Modgil, Adam M. Alessio, Michael D. Bindschadler, Kevin J. Little, David Rigie, Phillip A. Vargas and Patrick J. La Rivière:* Multi-dimensional sinogram restoration for myocardial blood flow estimation from dose-reduced dynamic CT .....p217.
11:30 - 11:50	*Michael T. Manhart, Andreas Fieselmann, Yu Deuerling-Zheng, Andreas K. Maier and Markus Kowarschik Dynamic*: Reconstruction with Statistical Ray Weighting for C-Arm CT Perfusion Imaging.....p221.

Wednesday, June 19

Oral Session: High Performance Computing (8:20 - 10:00 AM)

Session Chair: Marc Kachelrieß

8:20 - 8:40	*Timo Zinsser, Benjamin Keck*: Systematic Performance Optimization of Cone-Beam Back-Projection on the Kepler Architecture.....p225.
8:40 - 9:00	*Awen Autret, Julien Bert, Olivier Strauss and Dimitris Visvikis*: Fully 3D PET List-Mode reconstruction including an accurate detector modeling on GPU architecture.....p229.
9:00 - 9:20	*Matthias Baer and Marc Kachelrieß*: High Performance Parallel Beam and Perspective Cone-Beam Backprojection for CT Image Reconstruction on Pre–Production Intel R Xeon Phi.....p233.
9:20 - 9:40	*Eric Papenhausen, Ziyi Zheng, and Klaus Mueller*: Cloud X: A Platform as a Service for CT Reconstruction Research and Development.....p237.
9:40 - 10:00	*Jeffrey M. Rosen, Junjie Wu, Jeffrey A. Fessler, Thomas F. Wenisch*: Iterative Helical CT Reconstruction in the Cloud for Ten Dollars in Five Minutes.....p241.

Oral Session: Acquisition Geometries (10:30 - 11:50 AM)

Session Chair: Frédéric Noo

10:30 -10:50	*Zhicong Yu, Frédéric Noo, Günter Lauritsch, and Joachim Hornegger*: Extended volume image reconstruction using the Ellipse-Line-Ellipse trajectory for a C-arm system……p245.
10:50 - 11:10	*Brian E. Nett, Kai Zeng, and Jed D. Pack* : Image Reconstruction from an Axial Short Scan using a Katsevich type algorithm.....p249.
11:10 - 11:30	*Rolf Clackdoyle and Laurent Desbat*: Full Cone-Beam Consistency Conditions for Sources on a Plane.....p253.
11:30 - 11:50	*J. Webster Stayman and Jeffrey H. Siewerdsen*: Task-Based Trajectories in Iteratively Reconstructed Interventional Cone-Beam CT.....p257.

Oral Session: System Design and Calibration (16:00 - 17:40 PM)

Session Chair: Michael A. King

16:00-16:20	*Paul E. Kinahan, Chengeng Zeng, Larry A. Pierce II, Kalpana M. Kanal, Lawrence R. MacDonald*: Attenuation Correction Using a Single-View Mammogram for a 3D PET Scanner.....p261.
16:20-16:40	*Christina Debbeler, Nicole Maass, Matthias Elter, Frank Dennerlein, and Thorsten M. Buzug*: A New CT Rawdata Redundancy Measure applied to

	Automated Misalignment Correction.....….p264.
16:40-17:00	*Tobias Schon, Theobald Fuchs, Kilian Dremel, Christian Reuß*: A Translation-based Data Acquisition Method for Industrial Computed Tomography: Experimental Results.....p268.
17:00-17:20	*Abdelmoula Haboub, Alastair A. MacDowell, Stefano Marchesini, Diworth Y. Parkinson*: Coded Aperture Imaging for Fluorescent X-rays-Biomedical Applications.....….p272.
17:20-17:40	*Wojciech Zbijewski, J. Webster Stayman, Katsuyuki Taguchi, Erik Fredenberg, and Jeffrey H. Siewerdsen*: Volumetric Imaging with Sparse Arrays of Photon Counting Silicon Strip Detectors……...p276.

Poster Session (13:30 - 15:30 PM)

Session Chair: Benjamin M. W. Tsui

- *Hung Nien and Jeffrey A. Fessler*: Combining Augmented Lagrangian Method with *Ordered* Subsets for X-Ray CT Reconstruction.....….p280.
- *Madison G. McGaffin and Jeffrey A. Fessler*: Sparse shift-varying FIR preconditioners for fast volume denoising.....….p284.
- *Sajid Abbas, Taewon Lee, Hyekyun Chung, Jongduk Baek, and Seungryong Cho*: Effects of sparse sampling schemes on image quality in low-dose CT……p288.
- *Yanguang Lin, Bing Bai, Wentao Zhu, Ran Ren, Quanzheng Li, Richard M. Leahy*: Optimized MAP Reconstruction of H2-weighted Fourier Rebinned TOF PET…..……p292.
- *Ahmadreza Rezaei, Johan Nuyts, Michel Defrise*: The Effect of Motion on Joint Estimates of Activity and Attenuation from Time-of-Flight PET Data……...p296.
- *Jed D. Pack, Zhye Yin, Kai Zeng, Brian E. Nett*: Mitigating cone-beam artifacts in cardiac CT imaging for large cone-angle scans.....p300.
- *Stefan Sawall, Matthias Baer, Marcus Brehm, Michael Knaup, and Marc* Kachelrieß: Fast Computation of Projections from Triangulated Surfaces……..p304.
- Alexandre Bousse, Kjell Erlandsson, Stefano Pedemonte, Sebastien Ourselin, Simon R. Arridge, Brian F. Hutton: Angular Rebinning for Geometry Independent SPECT Reconstruction.....p308.
- *Daniil Kazantsev,* Sebastien *Ourselin, Brian F. Hutton, Simon R. Arridge*: A novel method of embedding additional information into tensor diffusion filtering as an application for multi-modal reconstruction in ET.....p312.
- *Hao Gao, X.* Sharon *Qi, and Daniel A. Low*: Imaging Quality Improvement and Dose Reduction on TomoTherapy via Tensor Framelet: Phantom Study……p316.
- *Radin A.* Nasirudin*, Kai Mei, Petar Penchev, Ernst J. Rummeny, Martin Fiebich, Peter B.*

Thursday, June 20

Oral Session: PET&SPECT (8:20 - 10:00 AM)

Session Chair: Michel Defrise

8:20 - 8:40	*Scott D. Metzler, Yusheng Li, Joel S. Karp, Samuel Matej*: Super-Resolution PET Using Stepping of a Deliberately Misaligned Bed.....p388.
8:40 - 9:00	*Mahmoud Abdalah, Rostyslav Boutchko, Debasis Mitra, Shi Chen, Grant T. Gullberg*: A Dynamic Regularization for Time Activity Curve Estimation in Dynamic Pinhole SPECT.....p392.
9:00 - 9:20	*Paul Dasari, Arda Könik, P. Hendrik Pretorius, Mohammed S. Shazeeb, W. Paul Segars, Karen L. Johnson and Michael A. King*: Evaluation of Bouc-Wen Model Corrected Respiratory Motion in Cardiac SPECT.....p396.
9:20 - 9:40	*Yong Long, Lishui Cheng, Xue Rui, Bruno De Man, Adam M. Alessio, Evren Asma and Paul E. Kinahan*: Analysis of Ultra-Low Dose CT Acquisition Protocol and Reconstruction Algorithm Combinations for PET Attenuation Correction.....p400.
9:40 - 10:00	*Hao Li, Georges El Fakhri, Quanzheng Li*: Direct MAP Estimation of Attenuation Sinogram using TOF PET Data and Anatomical Image.....p404.

Oral Session: Artifact Reduction and Postprocessing (10:30 - 11:50 AM)

Session Chair: Jiang Hsieh

10:30 -10:50	*Ruoqiao Zhang, Jean-Baptiste Thibault, Charles A. Bouman, Ken D. Sauer*: Soft Classification with Gaussian Mixture Model for Clinical Dual-Energy CT Reconstructions.....p408.
10:50 - 11:10	*Yanbo Zhang, Xuanqin Mou*: Metal artifact reduction based on the combined prior image.....p412.
11:10 - 11:30	*Dieter Hahn, Pierre Thibault, Andreas Fehringer, Martin Bech, Peter B. Noël and Franz Pfeiffer*: Bone artifact reduction in differential phase-contrast CT.p416.
11:30 - 11:50	*Johan Sunnegardh, Karl Stierstorfer*: A new method for windmill artifact reduction.....p420.

Oral Session: Regularization Methods (16:00 - 17:40 PM)

Session Chair: Jeffrey A. Fessler

16:00-16:20	*Hao Dang, Adam S. Wang, Zhe Zhao, Marc Sussman, Jeffrey H. Siewerdsen, J. Webster Stayman*: Joint Estimation of Deformation and Penalized-Likelihood CT Reconstruction Using Previously Acquired Images.p424.
16:20-16:40	*Kevin M. Brown, Thomas Koehler, Frank Bergner, Rolf Bippus, Bernhard Brendel, Stanislav Zabic, W. Clem Karl, Sarabjeet Singh, Atul Padole, and Synho Do*: Sparse Sampling for CT Dose Reduction.....p428.
16:40-17:00	*Katharina Schmitt, Harold Schöndube, Karl Stierstorfer, Joachim Hornegger, Frédéric Noo*: Challenges posed by statistical weights and data redundancies in iterative X-ray CT reconstruction.....p432.
17:00-17:20	*Tetsuya Kobayashi, Keishi Kitamura and Hiroyuki Kudo*: Anatomically-guided MAP Reconstruction of Partial-ring TOF PET Data Using Spots-on-Smooth Image Representation Model. .....p436.
17:20-17:40	*Junfeng Wu, Xuanqin Mou, Hengyong Yu, Ge Wang*: Statistical Interior Tomography via Dictionary Learning without Assuming an Object Support.p440.

Poster Session (13:30 - 15:30 PM)

Session Chair: Wenli Wang

- *Adrian A. Sanchez, Emil Y. Sidky, and Xiaochuan Pan*: Investigation of Template Structure for a Cone-Beam CT Signal Detection Task.....p444.
- *Guangzhi Cao, Brian E. Nett, Scott Hsieh, Jed D. Pack, and Jiang Hsieh*: Cone-beam Reconstruction and Evaluation for Helical Tomography……..p448.
- *Hao Zhang, Hao Han, Yan Liu, Hongbing Lu, Jianhua Ma, Jing Wang, and Zhengrong Liang*: Penalized weighted least-squares image reconstruction for low-dose CT using adaptive MRF coefficients predicted from normal-dose scan.....p452.
- *Yannan Jin, Geng Fu, Vladimir Lobastov, Peter M. Edic, Bruno De Man*: Dual-energy performance of x-ray CT with an energy-resolved photon-counting detector in comparison to x-ray CT with dual kVp.....p456.
- *Steven Bartolac and David Jaffray*: Practical Noise Correlation Modeling for Fluence Field Modulated Computed Tomography.....p461.
- *Synho Do, W. Clem Karl, Thomas Brady, Georges El Fahkri, and Rajiv Gupta*: A Non-Uniform Super-Resolution Compressive Sampling Method For Tomographic Imaging.....p465.

Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods

Donghwan Kim, Sathish Ramani, and Jeffrey A. Fessler

***Abstract*—Low-dose X-ray CT can reduce the risk of cancer to patients. However, it requires computationally expensive statistical image reconstruction methods for improved image quality. Iterative algorithms require long compute times, so we focus on algorithms that "converge" in few iterations. This paper proposes to apply ordered subsets (OS) methods to Nesterov's fast first-order methods for 3D X-ray CT problems. Nesterov's algorithms use previous iterates to provide momentum towards the optimum and thus achieve a fast convergence rate of $O(1/n^2)$, where n counts the number of iterations. We also propose to use separable quadratic surrogates (SQS) (with a non-uniform (NU) approach) in Nesterov's algorithms. We use a real patient helical CT scan to show that the proposed algorithms converge rapidly, and we investigate the behavior of OS methods in Nesterov's algorithms.**

I. Introduction

Based on the statistics of X-ray CT, we reconstruct a (non-negative) image $x \in \mathbb{R}_+^{N_p}$ from noisy measurements $y \in \mathbb{R}^{N_d}$ by minimizing a convex and continuously differentiable objective function $\Psi(x)$. This paper focuses on a penalized weighted least squares (PWLS) problem [1]:

$$\hat{x} = \arg\min_{x \succeq 0} \left\{ \Psi(x) \triangleq \frac{1}{2}||y - Ax||_W^2 + \beta R(x) \right\}, \quad (1)$$

where A is a projection operator, and the diagonal matrix W provides statistical weighting. $R(x)$ is a (edge-preserving) regularization function and β balances the data-fit term and $R(x)$. Due to the large scale of the problem (in 3D CT), iterative algorithms for minimizing $\Psi(x)$ require considerable compute time. Thus, the goal of this paper is to develop iterative algorithms that "converge" in fewer iterations.

This paper focuses on Nesterov's fast first-order algorithms [2], [3] that use previous iterates as momentum for additional acceleration towards the optimum. The former [2] uses two previous iterates as momentum, while the latter [3] uses all accumulated previous iterates. Both provide a fast convergence rate of $O(1/n^2)$ where n counts the number of iterations, whereas usual gradient-based methods have $O(1/n)$ convergence rate [4].

In our recent work [5], we combined ordered subsets (OS) methods [6], [7] with Nesterov's early work [2] that has been used to develop a fast iterative shrinkage-thresholding algorithm (FISTA) [4]. We also used a separable quadratic surrogates (SQS) method [7] (and a non-uniform approach [8]) in Nesterov's algorithm (in [5]). These combinations provided very promising results as they converged very rapidly even with relatively small number of subsets. (Using fewer subsets is preferable, as it decreases inexactness in OS methods and also reduces the overhead of computing the regularizer.) In addition, the overhead needed for proposed algorithms with OS-SQS is minimal as Nesterov's algorithms are simple. In this paper, we apply OS and (NU-)SQS methods to the more recent Nesterov's algorithm (2005) [3] and observe that this combination achieves as fast a convergence as the method in [5] but with improved stability.

We propose to use OS methods here as they can initially accelerate any gradient-based algorithms dramatically by approximating $\nabla\Psi(x)$ using only a subset of measurements. But, OS methods usually approach a limit-cycle looping around the optimum [6], [7]. (The more the subsets, the more the initial acceleration but with increased inexactness in the iterates.) However, the stability of OS methods in the proposed Nesterov's algorithms is unknown. Therefore, we experimentally investigated the behavior of OS in Nesterov's algorithms with respect to the number of subsets. We found that our newly proposed Nesterov's algorithm based on [3] with OS-SQS is more stable than the previous combination in [5].

In this paper, we propose to combine OS-SQS methods with Nesterov's fast first-order algorithms for X-ray CT image reconstruction. We first explain two of Nesterov's algorithms and illustrate their application to the X-ray CT problem in (1) with OS-SQS algorithms. Then we show the results for accelerated convergence of the two proposed algorithms using a real patient CT scan. We also discuss the stability of OS in Nesterov's algorithms.

II. Nesterov's algorithms

Nesterov published a fast first-order method using two previous iterates as a momentum for smooth functions[1] in [2], and it was extended later for non-smooth functions by Beck *et al.* [4], which is one of the state-of-the-art methods in image restoration. In [3], Nesterov also proposed new formulation of a fast first-order method using all previous iterates.

Both algorithms [2], [3] have been used widely for various optimization problems. They have also been used for X-ray CT reconstruction showing a noticeable acceleration [9], [10]. However, Nesterov's algorithms by themselves are not very

D. Kim, S. Ramani, and J. A. Fessler are with the Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail:kimdongh@umich.edu, sramani@umich.edu, fessler@umich.edu).

Supported in part by NIH grant R01-HL-098686 and equipment donations from Intel. Helical CT data provided by GE Healthcare.

[1] A smooth function $f(x)$ is continuously differentiable with Lipschitz continuous gradient L satisfying $||\nabla f(x) - \nabla f(z)|| \leq L||x - z||$ for all $x, z \in \mathbb{R}^{N_p}$.

attractive in CT, as the cost function $\Psi(x)$ in (1) has a large Lipschitz constant that slows down the convergence [11]. Here, we suggest new combinations of Nesterov's algorithms and OS-SQS methods that show very promising results.

We first review Nesterov's algorithms briefly. Both [2] and [3] begin by using an optimization transfer technique [12]. Nesterov uses a convex cost function $\Psi(x)$ that is continuously differentiable with Lipschitz constant L, which can be majorized at the nth iteration as:

$$\begin{aligned}\Psi(x) &\le \phi_L^{(n)}(x) \qquad (2)\\ &\triangleq \Psi(x^{(n)}) + \nabla\Psi(x^{(n)})'(x - x^{(n)}) + \frac{L}{2}||x - x^{(n)}||^2.\end{aligned}$$

The optimization transfer step minimizes the surrogate $\phi_L^{(n)}(x)$ at nth iteration:

$$x^{(n+1)} = \arg\min_{x \succeq 0} \phi_L^{(n)}(x) = \left[x^{(n)} - \frac{1}{L}\nabla\Psi(x^{(n)})\right]_+, \quad (3)$$

where $[\cdot]_+$ enforces a non-negativity constraint. Then the algorithm (3) is accelerated using previous iterates as shown in Figs. 1 and 2 [2], [3]. We use the choice of parameters suggested in [13] for the algorithm in Fig. 2, which provides faster convergence than the choice in [3].

Initialize $x^{(0)} = v^{(0)},\ t_0 = 1$
for $n = 0, 1, 2, \cdots$

$$t_{n+1} = \left(1 + \sqrt{1 + 4t_n^2}\right)/2$$
$$x^{(n+1)} = \left[z^{(n)} - \frac{1}{L}\nabla\Psi(z^{(n)})\right]_+$$
$$z^{(n+1)} = x^{(n+1)} + \frac{t_n - 1}{t_{n+1}}(x^{(n+1)} - x^{(n)})$$

Fig. 1. Nesterov's algorithm (1983) [2].

Initialize $x^{(0)} = v^{(0)} = z^{(0)},\ t_0 = 1$
for $n = 0, 1, 2, \cdots$

$$t_{n+1} = \left(1 + \sqrt{1 + 4t_n^2}\right)/2$$
$$x^{(n+1)} = \left[z^{(n)} - \frac{1}{L}\nabla\Psi(z^{(n)})\right]_+$$
$$v^{(n+1)} = \left[z^{(0)} - \frac{1}{L}\sum_{k=0}^{n} t_k \nabla\Psi(z^{(k)})\right]_+$$
$$z^{(n+1)} = \left(1 - \frac{1}{t_{n+1}}\right)x^{(n+1)} + \frac{1}{t_{n+1}}v^{(n+1)}$$

Fig. 2. Nesterov's algorithm (2005) [3].

The sequences $\left\{x^{(n)}\right\}$ generated by both algorithms are proven to have the following convergence rate [2], [3]:

$$\Psi(x^{(n)}) - \Psi(\hat{x}) \le O\left(\frac{L}{n^2}\right). \quad (4)$$

This is promising since ordinary optimization transfer in (3) provides only $O(1/n)$ rate [4]. However, the large Lipschitz constant L in CT problem causes slow convergence even with the $O(1/n^2)$ rate.

III. Proposed Nesterov's algorithms with ordered subsets

We suggest combining ordered subsets with Nesterov's fast first-order algorithms. Ordered subsets algorithms group projection views into M subsets evenly, and assume

$$\nabla\Psi(x) \approx M\nabla\Psi_0(x) \approx \cdots \approx M\nabla\Psi_{M-1}(x), \quad (5)$$

where we define the subset gradient:

$$\nabla\Psi_m(x) \triangleq A_m' W_m (A_m x - y_m) + \frac{\beta}{M}\nabla R(x) \quad (6)$$

for $m = 0, \cdots, M-1$. The matrices A_m, y_m and W_m are sub-matrices of A, y, and W corresponding to mth subset. We accelerate Nesterov's algorithms by replacing $\nabla\Psi(\cdot)$ in Figs. 1 and 2 with $M\nabla\Psi_m(\cdot)$. We count each mth sub-iteration as $1/M$ iteration, since $M\nabla\Psi_m(\cdot)$ requires roughly $1/M$ amount of computation of $\nabla\Psi(\cdot)$. Then we expect to have the following convergence rate in early iterations:

$$\Psi(x^{(n+\frac{m}{M})}) - \Psi(\hat{x}) \lessapprox O\left(\frac{L}{(nM+m)^2}\right). \quad (7)$$

This rate will not hold as the sequence $\left\{x^{(n+\frac{m}{M})}\right\}$ nears the optimum where the condition (5) fails.

Owing to the acceleration in proposed algorithms based on the M^2 effect of OS in (7), it is possible to use fewer subsets for better accuracy in OS. However, it is unknown how the inexactness in OS methods affect the behavior of the Nesterov's algorithms. (Ordinary OS methods are known to reach a limit-cycle looping around the optimum.) Therefore, we investigated OS algorithms with Nesterov's algorithms in Section IV, where we found that Nesterov's algorithm (2005) with OS methods is better stabilized than the earlier one.

For CT, it is computationally expensive to find the smallest possible Lipschitz constant L, and the backtracking line search scheme in [4] would be undesirably slow. Instead, we use a separable quadratic surrogate (SQS) method [7] for the optimization transfer step in (3), replacing $\phi_L^{(n)}$ in (2) by

$$\phi_{SQS}^{(n)}(x) \triangleq \Psi(x^{(n)}) + \nabla\Psi(x^{(n)})'(x - x^{(n)}) + \frac{1}{2}||x - x^{(n)}||_D^2, \quad (8)$$

where D is a diagonal matrix. The advantage of using SQS is that we can compute an exact surrogate $\phi_{SQS}^{(n)}(x)$ with modest computation. We can further accelerate the SQS-type algorithms by our recently proposed non-uniform approach [8].

We summarize the proposed algorithms, namely OS-SQS-Nes83 and OS-SQS-Nes05, in Figs. 3 and 4 that respectively combine OS-SQS with the two methods of Nesterov in Figs. 1 and 2.

IV. Results

We used a 3D helical X-ray CT data set of a human shoulder to show the acceleration of proposed algorithms. We computed the root mean square difference (RMSD) between the current and converged[2] image within the region-of-interest (ROI) in

[2]We generated an (almost) converged image by running 100 iterations of (convergent) NH-ABCD-SQS [8] followed by 2000 iterations of (convergent) SQS.

Initialize $x^{(0)} = v^{(0)}, t_0 = 1$

for $n = 0, 1, 2, \cdots$

for $m = 0, 1, \cdots, M-1$

$$t_{nM+m+1} = \left(1 + \sqrt{1 + 4t_{nM+m}^2}\right)/2$$

$$x^{(n+\frac{m+1}{M})} = \left[z^{(n+\frac{m}{M})} - D^{-1} M \nabla \Psi_m(z^{(n+\frac{m}{M})})\right]_+$$

$$z^{(n+\frac{m+1}{M})} = x^{(n+\frac{m+1}{M})} + \frac{t_{nM+m} - 1}{t_{nM+m+1}}(x^{(n+\frac{m+1}{M})} - x^{(n+\frac{m}{M})})$$

Fig. 3. Proposed Nesterov's algorithm (1983) with ordered subsets (OS-Nes83).

Initialize $x^{(0)} = v^{(0)} = z^{(0)}, t_0 = 1$

for $n = 0, 1, 2, \cdots$

for $m = 0, 1, \cdots, M-1$

$$t_{nM+m+1} = \left(1 + \sqrt{1 + 4t_{nM+m}^2}\right)/2$$

$$x^{(n+\frac{m+1}{M})} = \left[z^{(n+\frac{m}{M})} - D^{-1} M \nabla \Psi_m(z^{(n+\frac{m}{M})})\right]_+$$

$$v^{(n+\frac{m+1}{M})} = \left[z^{(0)} - D^{-1} \sum_{k=0}^{nM+m} t_k M \nabla \Psi_{(k)_M}(z^{(\frac{k}{M})})\right]_+$$

$$z^{(n+\frac{m+1}{M})} = \left(1 - \frac{1}{t_{nM+m+1}}\right) x^{(n+\frac{m+1}{M})} + \frac{1}{t_{nM+m+1}} v^{(n+\frac{m+1}{M})}$$

Fig. 4. Proposed Nesterov's algorithm (2005) with ordered subsets (OS-Nes05). The notation $(k)_M$ stands for $k \bmod M$.

Hounsfield Units (HU):

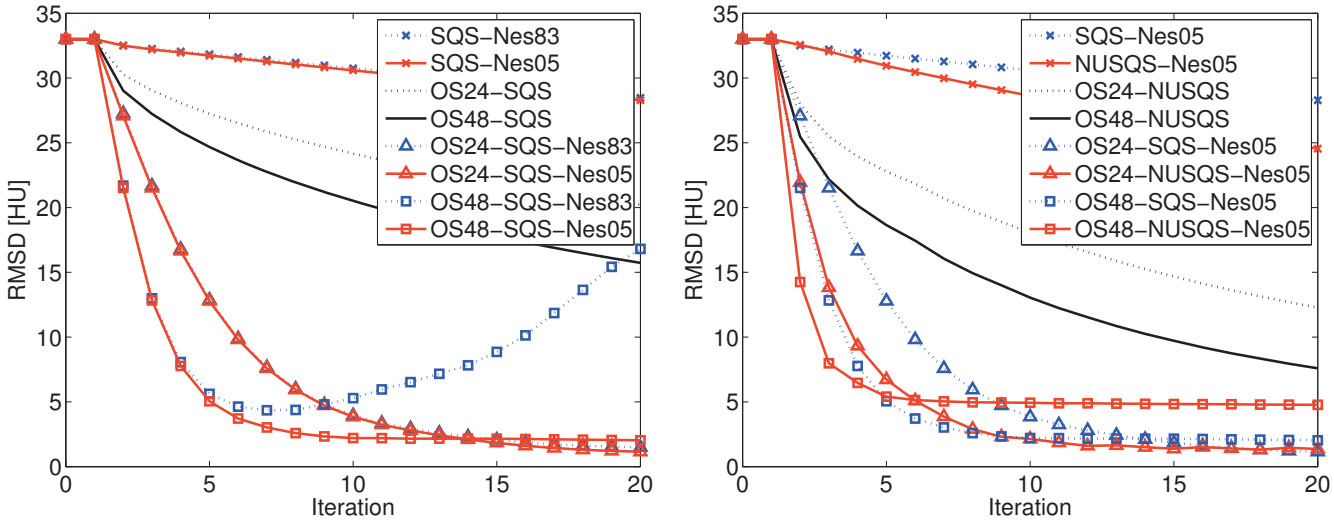
$$\text{RMSD} = \frac{||x_{\text{ROI}}^{(n)} - \hat{x}_{\text{ROI}}||_2}{\sqrt{N_{\text{p,ROI}}}} [\text{HU}] \tag{9}$$

versus iteration, to evaluate the convergence rate. In Fig. 5(a), we used different number of subsets such as 1, 24, and 48 subsets and observed that the ordered subsets highly accelerated both Nesterov's algorithms.

However, OS-Nes83 algorithms diverged when we used more than 40 subsets (as seen in the case of 48 subsets in Fig. 5(a)), while OS-Nes05 algorithm remained stable with more than 100 subsets. (Results not shown here.) Based on our observations, we believe that OS-Nes05 is more stable than OS-Nes83. We can intuitively understand this behavior, since OS-Nes05 method uses accumulated momentum that is less prone to local inexactness, while OS-Nes83 uses the difference between two previous iterates as momentum which may be very inaccurate in OS-type methods. However, we need theoretical justification to better understand the behavior of OS in Nesterov's algorithms, and we leave it as a future work.

We also combined a non-uniform (NU) approach [8] with OS-SQS-Nes05 to investigate the net resulting acceleration. In Fig. 5(b), we obtained some acceleration when including NU, but the algorithm reached a larger limit-cycle than the case without NU. OS-Nes05 with 24 subsets showed promising acceleration (with a slightly larger limit-cycle), but the algorithm with 48 subsets reached a quite large limit-cycle after initial acceleration. Further refinement of NU method is needed to reach a relatively small limit-cycle while achieving noticeable acceleration for large M.

Fig. 6 presents the initial filtered back projection (FBP) image $x^{(0)}$, the converged image $\hat{x}$, and reconstructed images at 12th iteration from four different algorithms for comparison. Both SQS-Nes05 and OS48-NUSQS at 12 iteration are still far from the converged image. The proposed algorithms OS48-SQS-Nes05 and OS24-NUSQS-Nes05 reach low RMSD level after 10 iterations in Fig. 5(b), and their reconstructed images at 12 iteration are very close to the converged image. The



(a) (b)

Fig. 5. Plots of RMSD in (9) versus iterations for various proposed Nesterov's algorithms with OS-NUSQS. (There are no changes in RMSD during the first iteration, since we count the precomputation of D as one iteration.)

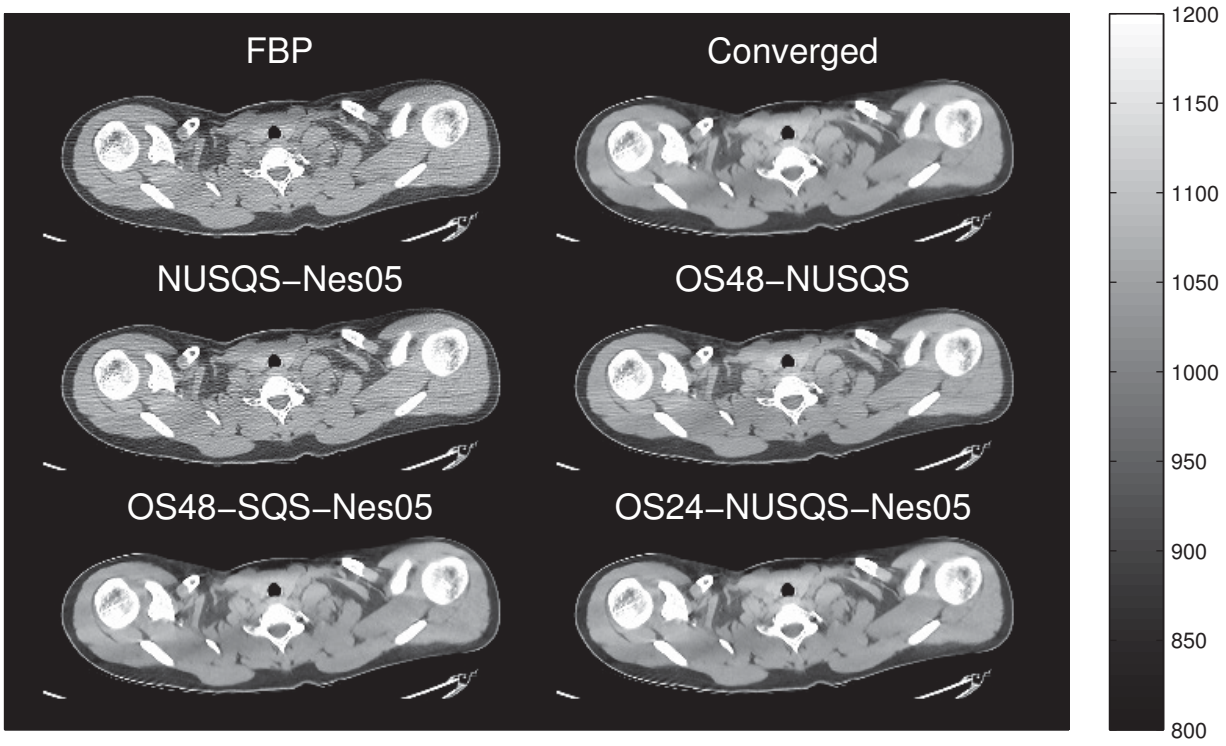


Fig. 6. Center slice of FBP image $x^{(0)}$, converged image $\hat{x}$, and reconstructed images at 12th iteration.

results confirm that the proposed combinations of OS and Nesterov's algorithms reach a decent image (close to $\hat{x}$) in few iterations.

V. Discussion

In this paper, we used a helical CT data set that corresponds to 984 projection views per turn with pitch 1.0. From the results, we were able to assess the behavior of OS-(NU)SQS-Nes empirically for this specific geometry. However, the number of subsets used for this geometry may not be optimal for other geometries. So, it is important to investigate the problem of selecting the appropriate number of subsets for a given geometry that would ensure fast convergence without encountering stability issues.

VI. Conclusion

We proposed two algorithms that combine Nesterov's methods with OS. The proposed algorithms provide dramatic acceleration in X-ray CT reconstruction with relatively small number of subsets. We found that the Nesterov's algorithm (2005) [3] is more stable with ordered subsets than the other choice [2] in our experiment. But, this should be examined on various other data sets, and we leave the theoretical justification as a future work.

Here, we investigated two specific methods [2], [3] for combining "momentum" terms with ordered subsets. There are many other possible ways to introduce momentum into OS methods and our future work aims at finding ways that are fast yet relatively stable for OS-type updates.

References

[1] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[2] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Math. Dokl.*, vol. 27, no. 2, pp. 372–76, 1983.

[3] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–52, May 2005.

[4] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[5] D. Kim, S. Ramani, and J. A. Fessler, "Ordered subsets with momentum for accelerated X-ray CT image reconstruction," in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2013, To appear.

[6] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–9, Dec. 1994.

[7] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[8] D. Kim and J. A. Fessler, "Parallelizable algorithms for X-ray CT image reconstruction with spatially non-uniform updates," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 33–6.

[9] K. Choi, J. Wang, L. Zhu, T-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Med. Phys.*, vol. 37, no. 9, pp. 5113–25, Nov. 2010.

[10] J. H. Jorgensen, T. L. Jensen, P. C. Hansen, S. H. Jensen, E. Y. Sidky, and X. Pan, "Accelerated gradient methods for total-variation-based CT image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011, pp. 435–8.

[11] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–88, Mar. 2012.

[12] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Im. Proc.*, vol. 16, no. 10, pp. 2411–22, Oct. 2007.

[13] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," 2008.

Extending the Dynamic Range of Flat Detectors in CBCT using a Compressed–Sensing–Based Multi–Exposure Technique

Ludwig Ritschl, Michael Knaup and Marc Kachelrieß

Abstract— **The limited dynamic range of flat detectors is one of the main reasons for reduced soft tissue visibility in flat detector–based CBCT. Here the main limiting size is the detector's background electronic noise which usually reduces the effective bitdepth of the detector to a range of about 7-9 bit. This limited dynamic results in a trade–off between pixel saturation in air and weakly absorbing areas and an insufficient grayscale resolution of strong absorbing areas. While the former should be avoided to enable for quantitative reconstruction, the latter leads to a loss of low-contrast visibility inside the object. In this study we propose a multi–exposure technique which yields two datasets. The first dataset is acquired at a very low dose level and using only a small number of projections and guarantees that the detector is not saturated. The second dataset is fully sampled at the standard dose level which is required for visualizing certain soft tissue structures. The final volume will be reconstructed using both datasets simultaniosly using a statistically–weighted total variation–constrained iterative reconstruction algorithm. To evaluate the method a simulation study is performed.**

I. Introduction

Flat detector–based cone–beam computed tomography (CBCT) is a widely used imaging modality. Typical devices in the medical field are C-arms, dental CT scanners and onboard imaging systems for image guided radiation therapy. Compared to clinical CT scanners there are several factors leading to reduced soft tissue visibility. Beside the lower photon absorption efficiency of flat detectors and scattered radiation the limited dynamic range of flat detectors is one of the main sources. In typical thin film transistor (TFT) flat detectors, as they are widely used, the background electronic noise of the detector can be assumed to be the limiting size in terms of grayscale resolution [1, 2]. That means that for a certain x–ray intensity I at the detector the electronic noise σ_{e^-} is larger than the Poisson–distributed quantum noise $\sigma_q = \sqrt{I}$. This intensity value will be denoted with I_{lim} and is a characteristic size for a flat detector with fixed amplification settings. Now the electronic noise limited dynamic of the detector can be calculated as

$$\text{Dynamic Range} = I_0/I_{\text{lim}}.$$

This limitation automatically leads to the fact that the smallest detectable intensity difference ΔI at intensity levels $I < I_{\text{lim}}$ is $\Delta I > \sigma_{e^-}$. Regarding the exponential attenuation of x–rays

$$I = I_0\, e^{-p},$$

with p being the attenuation value, one can see that for a fixed entrance dose I_0 the minimal detectable intensity difference ΔI is correlated to a minimal detectable difference of attenuation values Δp.

In the following we define the gain factor g in such a way that that for $g = 1$ and a given entrance intensity I_0 the detector is not saturated and it's full dynamic range is used. This would be a standard setting to avoid detector saturation which enables for correct CT image reconstruction. A value of $g = 2$ would mean that the detector is saturated at $I = 0.5\ I_0 = I_0/g$. To realize the gain g we assume an ideal analog amplification of the entrance signal in the detector which does not increase the electronic noise with the same factor. Now, the electronic noise limited dose level can be calculated by regarding $\sigma_{e^-} \geq g\sqrt{I}$, which automatically laeds to

$$I_{\text{lim}}^{g=0.5} = 0.25\ I_{\text{lim}}^{g=1.0}$$

in the mentioned case. Using this fact one can define acqusition settings combing the entrance dose I_0 and the corresponding gain g to make sure that the desired minimal visible attenuation difference Δp is limited by quantum noise. In most real world applications this leads to values $g > 1$ which results in object truncation, which makes a quantitatively correct reconstruction impossible (Figure 1). In literature there are different ways proposed to handle this problem. One way is using different gain factors during detector read out and combining the acquired images [1]. This method leads to a reduced detector resolution or reduced frame rate. [1]. Another approach is the use of so–called bowtie filters, which assume a circular motion around an object in the center of the field of view. Now, rays at the border of the object are prefiltered and attenuated to levels below I_0/g. This approach limits the applications to strictly centered objects.

In this paper we propopse a new approach to handle this problem. The main idea is to acquire two datasets, one at the optimum level of g, which is required for soft tissue visualization, the other one at a reduced entrance dose I_0/g, to make sure, that the detector is not saturated. Both datasets can be acquired during one scan using an AEC (automatic exposure control) where the pulselength of single projections can be varied.

II. Method

A. Multiple exposure–based data acquisition

The acquisition of the datasets p_1 with intensity I_0 and p_2 with intensity I_0/g would increase the radiation dose by a factor of $1 + 1/g$. To avoid a significant increase of radiation dose we propose a very sparse sampling in angular direction of the dataset p_2. In this study we used an undersampling factor of 20 regarding the required full sampling number of projections.

Dr. Ludwig Ritschl: Ziehm Imaging, Donaustr. 31, 90451 Nürnberg, Germany.

Prof. Dr. Marc Kachelrieß, Dr. Michael Knaup: Medical Physics in Radiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

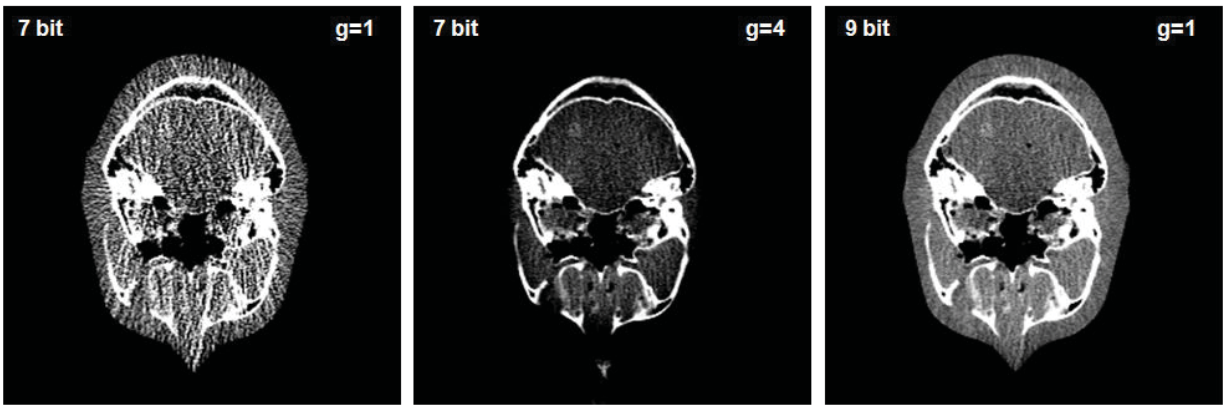


Fig. 1. The left image shows the simulated dataset using a detector dynamic range of 7 bit. The amplification is $g = 1$, so that there is no saturation in air. The poor soft tissue contrast visibility due to the insufficient detector dynamic is clearly visible. The second image is simulated with the same dynamic range but an amplification of $g = 4$, which leads to a better soft tissue visibility but also to truncation artifacts and an HU shift due to saturated pixel values on the detector. The right image shows the simulation at 9 bit dynamics, which is enough dynamic range to avoid saturation and loss of soft tissue visibility. All images are simulated at the same dose level. All images are windowed C/W = 0 HU / 1000 HU.

Both datasets are reconstructed using a total variation–based iterative reconstruction. Additionally a statistical weight is introduced into the reconstruction framework to account for the different photon statistics of both datasets.

B. TV–constrained reconstruction

To solve the reconstruction problem we aim to minimize the following cost function:

$$\begin{aligned} &\min ||\nabla f(r)||_1 \text{ subject to} \\ &((\mathsf{R}_1 f(r) - p_1)^T W (\mathsf{R}_1 f(r) - p_1) \\ &+ (\mathsf{R}_2 f(r) - p_2)^T W (\mathsf{R}_2 f(r) - p_2)) < \varepsilon. \end{aligned} \tag{1}$$

Here R_1 and R_2 denote the x–ray transform mapping the image $f(r)$ on the acquired raw data p_1 and p_2. W is a diagonal matrix with the entries $W_{i,i} = \sigma_q^{-2}$. The first term in equation 1 enforces a solution with the minimal total variation as described in references [3–5]. To find a solution of equation an approximative method, which has been developed for total variation–constrained image reconstruction, is used. For a more detailed description we refer the reader to reference [5].

III. Simulation

To simulate the effects of limited detector dynamics we assume a monochromatic 70 keV x–ray spectrum and perform the simulations and reconstructions in 2D parallel geometry with 512 projection angles and 512 rays per projection. Each ray in parallel geometry is parametrized by the two parameters ϑ and ξ such that $x\cos\vartheta + y\sin\vartheta = \xi$ is the line of integration in the x–y–plane.

The dataset used in our simulation is a CBCT of high quality which was used as reference. Raw data for the simulations were generated by forward projecting this dataset.

The line integrals obtained from the simulations are denoted as $p(\vartheta,\xi)$. The ideal rawdata $p(\vartheta,\xi)$ will be deteriorated by adding noise and simulating overexposure and discretization of AD converters, as explained in the following subsections.

A. Relative Intensities

The simulated line integrals $p(\vartheta,\xi)$ are converted into relative intensities as

$$q_0(\vartheta,\xi) = e^{-p(\vartheta,\xi)}. \tag{2}$$

Since these and all following manipulations are done detector pixel–wise we will drop the dependency on ϑ and ξ in the following.

B. Quantum Noise

We now add quantum noise as follows:

$$q_2 = q_1 + \mathcal{N}\sqrt{q_1/I_0}. \tag{3}$$

Here, $\mathcal{N}$ is a normal–distributed random number with mean 0 and standard deviation 1. I_0 is the number of detected quanta if no object was in the x–ray path and if the detector had no saturation limit. For this study, we chose a fixed $I_0 = 1.0 \times 10^6$ for all simulations.

C. Gain Factor

The previously defined gain is simulated by multiplying the the noisy intensity data q_2 with the gain factor g.

$$q_3 = g\, q_2. \tag{4}$$

D. Electronic Noise

This step adds electronic noise by performing

$$q_4 = q_3 + \mathcal{U} 2^{-b_e}. \tag{5}$$

Here, $\mathcal{U}$ is an uniformly distributed random number in the interval $[-\frac{1}{2}, \frac{1}{2}]$. b_e is the effective number of significant detector bits induced by electronic noise. Note that b_e will be a non–integral number in general although we restrict to integral numbers here.

E. AD Conversion and Saturation

Up to now, we performed all calculations in double precision which can be considered as an infinite precision compared to

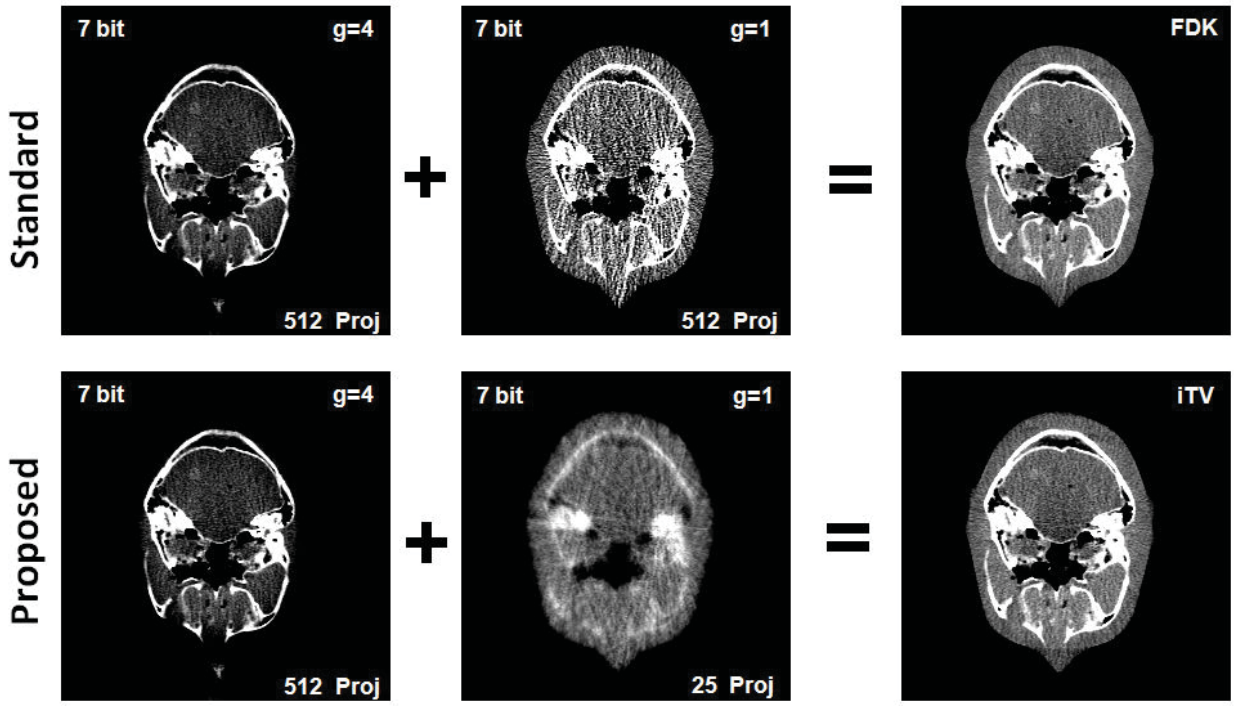


Fig. 2. The upper row shows a straight–forward way to extend the detectors dynamic range by combining the fully sampled datasets p_1 and p_2 in the raw data domain followed by standard FDK reconstruction. The second row shows the proposed method, where the dataset p_2 is sampled very sparse with only 25 projections. All images are windowed C/W = 0 HU / 1000 HU.

the actual detector resolution. In this step, we assume that the AD converter has a limited precision of b_d significant bits, i.e. it can deliver only 2^{b_d} distinct values. Note that usual detectors hold $b_d > b_e$, i.e. the detector dynamic range is usually limited by electornic noise, not by AD conversion. In this study, however, we set $b_d = b_e \equiv b$ for convenience. It figured out that for $8 \leq b_e \leq b_d$ the actual value of b_d (i.e. discretization) does not significantly affect the image quality.

We simulate discretization and saturation by the following formula:

$$q_5 = \varepsilon \vee \frac{\lfloor q_4(2^b - 1) + \frac{1}{2} \rfloor}{2^b - 1} \wedge 1. \qquad (6)$$

The maximum function clips to the small positive $\varepsilon = \frac{1}{2}(2^b - 1)^{-1}$ to avoid q_5 becomes zero (which would make trouble when taking the log in the next step). The minimum function clips the digitized value to the maximum relative intensity 1 which corresponds to a saturation at $I_{\max} = I_0/g$ quanta.

IV. Results

The simulation was performed combining different bit scales and gain factors of the detector. Note that the bit size describes the electronic noise limited dynamic of the detector. In Figure 2 two methods of detector dynamic extension are shown. The first row shows a straight–forward method where the two fully sampled datasets p_1 and p_2 are combined in the raw data domain and reconstructed using filtered backprojection. This method leads to an effective increase of the dynamic range from 7 bits to 9 bits as demonstrated in reference [2]. The increase of radiation dose in this case is 25%. The second row in Figure 2 shows the reconstruction using the proposed method. Here the increase of radiation dose is 1.25% which can be easily compenstaed by reducing the exposure dose of dataset p_1 for this amount. Comparing the result to the simulation at 9 bit dynamic range (Figure 1) one can see that the image quality is comparable to the quality of a dataset using a more sensitive detector. The full abstract will also show experimental CBCT data including quantitative evaluations of the results presented here.

V. Discussion

Extending the detector dynamic range in flat detector–based CBCT is one crucial step towards quantitative soft tissue visualization on CBCT systems. The method proposed in this paper shows an efficient way to do this. Compared to hardware–based methods (bowtie, multi gain), the method is extremely flexible and can be adapted to a wide range of object sizes and dose settings.

Acknowledgments

The high–speed image reconstruction software RayConStruct–IR was provided by RayConStruct GmbH, Nürnberg, Germany. We thank the Intel Cooperation and Fujitsu Technology Solutions GmbH for providing their latest multicore hardware.

References

[1] P. G. Roos, R. E. Colbeth, I. Mollov, P. Munro, J. Pavkovich, E. J. Seppi, E. G. Shapiro, C. A. Tognina, G. F. Virshup, J. M. Yu, G. Zentai, W. Kaissl, E. Matsinos, J. Richters, and H. Riem, "Multiple gain ranging readout method to extend the dynamic range of amorphous silicon flat panel imagers", *SPIE Medical Imaging Proc.*, vol. 5368, pp. 139–149, 2004.

[2] M. Knaup, L. Ritschl, M. Kachelrieß, "Low–Contrast Visibility in Flat Detector CT: A Simulation Study", *2nd CT Meeting*, 2012.

[3] E. Y. Sidky , C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Journal of X-Ray Science and Technology*, vol. 14, pp. 119-139, 2006.

[4] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by total variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777-4807, 2008.

[5] L. Ritschl, F. Bergner, C. Fleischmann and M. Kachelrieß. "Improved total variation–based CT image reconstruction applied to clinical data,"*Phys. Med. Biol.*, vol 56, pp. 1545-1561, 2011.

Efficient Compressed Sensing Statistical X-Ray/ CT Reconstruction from Fewer Measurements

Il Yong Chun and Thomas M. Talavage

***Abstract*—High quality (high resolution, contrast, and SNR) reconstructed x-ray/CT images can be achieved by emitting many photons with relatively high energy. Practical factors require a limited number of beams to reduce radiation dose, and minimal scanning time is desired to reduce motion artifacts. Compressed sensing can be a practical solution that accomplishes both goals with high image quality. However, perfect image reconstruction is non-achievable for limited measurements by l_1-norm convex minimization. An alternate solution is $l_{\mathrm{p}}(0<\mathrm{p}<1)$-quasi-norm non-convex minimization, an NP-hard problem. To solve this practically, a $l_{\mathrm{p},\varepsilon}(0<p<1, 0<\varepsilon\ll 0)$ penalty is introduced as an approximation, and an efficient reweighted l_1 minimization algorithm is proposed incorporating majorization-minimization, split-Bregman, and noise statistics. Moreover, a combined diagonal-symmetric preconditioner is proposed to resolve slow convergence caused by non-uniform Poisson noise variance. Simulation results exhibit almost perfect image reconstruction from limited measurements with rapid convergence.**

I. Introduction

Compressed Sensing (CS) can be applied to reconstruct a target image from few linear measurements that comprise a sparse subset of the full measurement. Assuming a system matrix $\mathbf{A}\in\mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ with $\mathrm{m}\leq\mathrm{n}$, a sparsifying transform $\boldsymbol{\psi}\in\mathbb{R}^{\mathrm{n}\times\mathrm{n}}$, and observations $\mathbf{y}\in\mathbb{R}^{\mathrm{m}}$, the most sparse solution $\mathbf{x}\in\mathbb{R}^{\mathrm{n}}$ of $\mathbf{Ax}=\mathbf{y}$ is ideally reconstructed using the cardinality of the vector, $\|\blacksquare\|_0$, as

$$\mathbf{x}^*=\underset{\mathbf{x}}{\operatorname{argmin}}\ \|\boldsymbol{\psi}\mathbf{x}\|_0,\ \text{s.t. }\mathbf{Ax}=\mathbf{y}. \tag{1}$$

Because it remains strongly NP-hard [22], (1) is typically solved using l_1-norm minimization when $\mathbf{A}$ satisfies several conditions including the restricted isometry property (RIP) [1], mutual coherence (MC) [2], and the null space property [2]. In practice, however, these conditions are too strong. Therefore, weaker sufficient conditions in terms of MC [20] and RIP [3] were developed to exactly recover sparse vectors by $l_{\mathrm{p}}(0<\mathrm{p}<1)$ -quasi-norm minimization. Specifically, [21] presents another version of the weaker sufficient condition in [3] and deals with the more realistic case of noisy $\mathbf{y}$:

$$\mathbf{x}^*=\underset{\mathbf{x}}{\operatorname{argmin}}\ \|\boldsymbol{\psi}\mathbf{x}\|_{\mathrm{p}}^{\mathrm{p}},\ \text{s.t. }\|\mathbf{Ax}-\mathbf{y}\|_2^2<\delta. \tag{2}$$

Non-convex minimization using an $l_{\mathrm{p}}(0<\mathrm{p}<1)$-quasi-norm has been shown to recover sparse signals from fewer linear measurements than are required for l_1 -norm convex minimization [3], [21], [23]. Although this remains a strongly NP-hard problem [5], it has been demonstrated in [3]-[5] that a local minimum can be computed in tractable time.

Using the discrete gradient transform as a sparsifying operator, CS theory has been successfully applied to x-ray/CT image reconstruction from limited numbers of measurements [6]-[8], but these solutions have not accounted for physical constraints imposed on the measurements by the actual device. Based on [9], the combination of x-ray quantum noise and detector electronic noise can be modeled as a compound additive Poisson-Gaussian random variable. Note however, that the Poisson noise variance is non-uniform, resulting in the expectation of slow convergence for iterative methods.

In this work, we use a $l_{\mathrm{p},\varepsilon}(0<\mathrm{p}<1,0<\varepsilon\ll 1)$ -penalty function to approximate $l_{\mathrm{p}}(0<\mathrm{p}<1)$ and propose an efficient p- and ε- dependent reweighted l_1 minimization algorithm based on the combination of majorization-minimization (MM) and the split-Bregman (SB) method, and incorporation of noise statistics. The reweighted l_1 minimization is known to recover sparse signals with lower error than a reweighted l_2-norm minimization algorithm (e.g., FOCUSS) [10], and the SB method is known to exhibit rapid and efficient convergence, especially when using an l_1-norm [11]. In contrast with the approach of splitting the non-uniform noise variance [12], we attacked the shift-variant component directly based on SB with a proposed combined diagonal-symmetric preconditioner based on the discrete cosine transform—dDCT—to decrease the number of splitting variables requiring update and to maintain rapid minimization convergence. The dDCT preconditioner exhibits faster convergence than any of diagonal, circulant (FFT, fast fourier transform), or combined diagonal-circulant (dFFT) [13] preconditioners.

II. Efficient Statistical Compressed Sensing X-Ray/CT Reconstruction with Non-Convex Penalty Function

A. Problem Formulation with $l_{p,\varepsilon}(0<p<1,0<\varepsilon\ll 1)$ from Majorization-Minimization to Compound Additive Poisson-Gaussian Noise

First, the non-convex minimization problem (2) is transformed into a reweighted l_1 convex minimization problem. Noting $\|\mathbf{x}\|_{\mathrm{p}}^{\mathrm{p}}=\sum_{\mathrm{i}=1}^{\mathrm{n}}|\mathrm{x_i}|^{\mathrm{p}}=\lim_{\varepsilon\to 0}\sum_{\mathrm{i}=1}^{\mathrm{n}}(|\mathrm{x_i}|+\varepsilon)^{\mathrm{p}}$, we can approximate the $l_p(0<\mathrm{p}<1)$-quasi-norm as

$$\|\mathbf{x}\|_{\mathrm{p}}^{\mathrm{p}}\approx l_{\mathrm{p},\varepsilon}(\mathbf{x})=\sum_{\mathrm{i}=1}^{\mathrm{n}}(|\mathrm{x_i}|+\varepsilon)^{\mathrm{p}}, \tag{3}$$

taking advantage of Lipschitz continuity. By (3), (2) becomes:

$$\mathbf{x}^*=\underset{\mathbf{x}}{\operatorname{argmin}}\ l_{\mathrm{p},\varepsilon}(\boldsymbol{\psi}\mathbf{x}),\ \text{s.t. }\|\mathbf{Ax}-\mathbf{y}\|_2^2<\delta. \tag{4}$$

School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. Email: chuni@purdue.edu, tmt@purdue.edu.

Given that any concave function is majorized by the tangent line, the following inequality holds for the $l_p (0 < p < 1)$-quasi-norm by linearization around x_i':

$$|x_i|^p \le |x_i'|^p + p|x_i'|^{p-1}(|x_i| - |x_i'|), \quad (5)$$

where $i = 1, \dots, n$. Because $l_{p,\varepsilon}(\mathbf{x})$ is separable, we can construct the following majorization function:

$$l_{p,\varepsilon}(\mathbf{x}) \le l_{p,\varepsilon}(\mathbf{x}') + \nabla l_{p,\varepsilon}(\mathbf{x}') \cdot (|\mathbf{x}| - |\mathbf{x}'|), \quad (6)$$

where $\nabla l_{p,\varepsilon}(\mathbf{x}') = [p(|x_1| + \varepsilon)^{p-1}, \quad \cdots \quad, p(|x_n| + \varepsilon)^{p-1}]^T$, $\cdot$ denotes a dot product, and $|\blacksquare|$ indicates the element-wise absolute value. Minimizing the majorization function with an iterative scheme improves the estimated solution of (4):

$$\mathbf{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \{ l_{p,\varepsilon}(\boldsymbol{\psi}\mathbf{x}^{(k)}) + \nabla l_{p,\varepsilon}(\boldsymbol{\psi}\mathbf{x}^{(k)}) \cdot (|\boldsymbol{\psi}\mathbf{x}| - |\boldsymbol{\psi}\mathbf{x}^{(k)}|) \}, \text{s.t.} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 < \delta, \quad (7)$$

which can be restated as the following convex MM algorithm:

$$\mathbf{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \|\mathbf{W}^{(k)}\boldsymbol{\psi}\mathbf{x}\|_1, \text{ s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 < \delta, \quad (8)$$

where $\mathbf{W}^{(k)} = diag(\mathbf{w}^{(k)})$, $w_i^{(k)} = p\left(\left|[\boldsymbol{\psi}\mathbf{x}^{(k)}]_i\right| + \varepsilon\right)^{p-1}$, $i = 1, \dots, n$, and $diag(\blacksquare)$ denotes conversion of a vector into a diagonal matrix. Based on the general theory described in [14], [15], $l_{p,\varepsilon}(\mathbf{x}^{(k)})$ converges to a local minimum of $l_{p,\varepsilon}(\mathbf{x})$.

The object linear attenuation coefficient image $\mathbf{x} \in \mathbb{R}^n$ is reconstructed with the following transformed unconstrained problem derived from the constrained problem (8) (from optimization theory, (8) and (9) are equivalent from the perspective that the two share the same solution):

$$\mathbf{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \left\{ \|\mathbf{W}^{(k)}\boldsymbol{\psi}\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \right\}, \quad (9)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is now the sparse forward projection matrix, and $\mathbf{y} \in \mathbb{R}^m$ is the measurement where $y_i = ln(\rho_0/\rho_i)$ with incident photon count ρ_0 and photon count measurement ρ_i. In practice, the ρ_0 may vary with projection, but we assume that ρ_0 is constant. Note that because the projection angles and beams are not randomly spaced, they result in the uniform distribution of energy over the image and a structure which will guarantee rapid convergence.

Although this method works well with Gaussian noise, the quadratic term becomes problematic in the presence of Poisson noise, such as in x-ray/CT. Because the variance of the noise in a Poisson model is proportional to the signal intensity, over-fitting and over-smoothing problems can result in high- and low-intensity regions, correspondingly. A weighting matrix $\boldsymbol{\Lambda}$ is used to resolve this problem:

$$\mathbf{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \left\{ \|\mathbf{W}^{(k)}\boldsymbol{\psi}\mathbf{x}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\boldsymbol{\Lambda}}^2 \right\}. \quad (10)$$

Modeling ρ_i as a sum of Poisson random variables with mean $\bar{\rho}_i$, and the electronic readout noise as $\mathcal{N}(0, \sigma^2)$, then the variance of y_i can be derived to be $\sigma_{y_i}^2 = (\bar{\rho}_i + \sigma^2)/\bar{\rho}_i^2$, and the diagonal elements of $\boldsymbol{\Lambda} = diag(\boldsymbol{\lambda})$ are given by

$$\lambda_i = \frac{\rho_i^2}{\rho_i + \sigma^2}, \quad (11)$$

where $i = 1, \dots, m$, because ρ_i is unbiased estimation of $\bar{\rho}_i$ [9].

B. Reweighted l_1-Norm Minimization by Split-Bregman

The l_1-regularized problem (10) may be reduced to a sequence of unconstrained problems with a Bregman update:

$$\left(\mathbf{x}^{(k+1)}, \mathbf{d}^{(k+1)}\right) = \underset{\mathbf{x},\mathbf{d}}{\operatorname{argmin}} \left\{ \|\mathbf{W}^{(k)}\mathbf{d}\|_1 + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\boldsymbol{\Lambda}}^2 + \frac{\gamma}{2}\|\mathbf{d} - \boldsymbol{\psi}\mathbf{x} - \mathbf{b}^{(k)}\|_2^2 \right\} \quad (12)$$

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \boldsymbol{\psi}\mathbf{x}^{(k+1)} - \mathbf{d}^{(k+1)}. \quad (13)$$

To solve (12) efficiently, we apply the SB technique to decompose the l_1 and l_2 components into two sub-problems, and solve by iterative minimization with respect to x and d:

$$\mathbf{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \left\{ \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\boldsymbol{\Lambda}}^2 + \frac{\gamma}{2}\|\mathbf{d}^{(k)} - \boldsymbol{\psi}\mathbf{x} - \mathbf{b}^{(k)}\|_2^2 \right\} \quad (14)$$

$$\mathbf{d}^{(k+1)} = \underset{\mathbf{d}}{\operatorname{argmin}} \left\{ \|\mathbf{W}^{(k)}\mathbf{d}\|_1 + \frac{\gamma}{2}\|\mathbf{d} - \boldsymbol{\psi}\mathbf{x}^{(k+1)} - \mathbf{b}^{(k)}\|_2^2 \right\}. \quad (15)$$

Note that (15) can be solved efficiently by an element-wise soft-shrinkage operator:

$$d_i^{(k+1)} = softshrink\left([\boldsymbol{\psi}\mathbf{x}^{(k+1)}]_i + b_i^{(k)}, w_i^{(k)}/\gamma\right), \quad (16)$$

where $softshrink(x, \alpha) = (x/|x|)\max(|x| - \alpha, 0)$ and $i = 1, \dots, n$.

Therefore, the total reconstruction time depends on the computational cost to solve (14), analytically as follows:

$$\mathbf{x}^{(k+1)} = inv(\mathbf{K})\mathbf{z}, \quad (17)$$

where $\mathbf{K} = \mu\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A} + \gamma\boldsymbol{\psi}^T\boldsymbol{\psi}$, $\mathbf{z} = \mu\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{y} + \gamma\boldsymbol{\psi}^T(\mathbf{d}^{(k)} - \mathbf{b}^{(k)})$, and $inv(\blacksquare)$ denotes a linear system solver to obtain $\mathbf{x}^{(k+1)}$ from $\mathbf{K}\mathbf{x}^{(k+1)} = \mathbf{z}$. If we choose an orthogonal sparsifying transform for $\boldsymbol{\psi}$, then $\mathbf{K} = \mu\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A} + \gamma\mathbf{I}$.

Input: Measurement y, $\mu > 0$, $\gamma > 0$, and $0 < p, \varepsilon \ll 1$
Pre-compute: dDCT preconditioner $\mathbf{M} = \mathbf{G}^{-1}\mathbf{C}^T\boldsymbol{\Omega}^{-1}\mathbf{C}\mathbf{G}^{-1}$ for $\mathbf{K} = \mu\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A} + \gamma\boldsymbol{\psi}^T\boldsymbol{\psi}$
Initialize: $\mathbf{x}^{(0)} = inv(\mathbf{A})\mathbf{y}$ by CG, $\mathbf{d}^{(0)} = \mathbf{b}^{(0)} = \mathbf{0}$, and $\mathbf{W}^{(0)} = \mathbf{I}$
While $tol(k) > \epsilon^2$ **do**
$\mathbf{x}^{(k+1)} = inv(\mathbf{K})\left(\mu\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{y} + \gamma\boldsymbol{\psi}^T(\mathbf{d}^{(k)} - \mathbf{b}^{(k)})\right)$ by PCG($\mathbf{M}$)
$\mathbf{d}^{(k+1)} = softshrink\left(\boldsymbol{\psi}\mathbf{x}^{(k+1)} + \mathbf{b}^{(k)}, (1/\gamma)\mathbf{W}^{(k)}\right)$
$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \boldsymbol{\psi}\mathbf{x}^{(k+1)} - \mathbf{d}^{(k+1)}$
$\mathbf{W}^{(k+1)} = diag(p(|\boldsymbol{\psi}\mathbf{x}^{(k+1)}| + \varepsilon)^{p-1})$
end
Output: Reconstructed image $\mathbf{x}$

Fig. 1. Efficient reweighted l_1-norm minimization algorithm using SB method for statistical CS X-ray/CT reconstruction.

C. A Combined Diagonal/DCT Preconditioning Methods

Because $\mathbf{K}$ is large, it is prohibitive to store and direct inversion is impractical. One of the well-known iterative methods to practically solve (17) when $\mathbf{K}$ is symmetric positive definite, is conjugate gradient (CG). However, CG converges slowly and an accurate preconditioner should be chosen for acceleration. For rapid convergence, (17) can be solved using preconditioned CG (PCG). We can intuitively consider the FFT-based preconditoner, because $\mathbf{A}^T\mathbf{A}$ has approximately Toeplitz-block-Toeplitz (TBT) structure. However, due to the non-uniform elements of $\boldsymbol{\Lambda}$, this is sub-optimal, and a different approach is warranted. Given $\mathbf{A}^T\mathbf{A}$ is only an approximate TBT matrix and that the point spread function (PSF) of $\mathbf{A}^T\mathbf{A}\delta_{(0,0)}$ (where $\mathbf{A}^T\mathbf{A}\delta_{(0,0)}$ means that a single pixel at the center of the image domain is projected and then back-projected) will not be perfectly symmetric, a DCT-based preconditioner with the

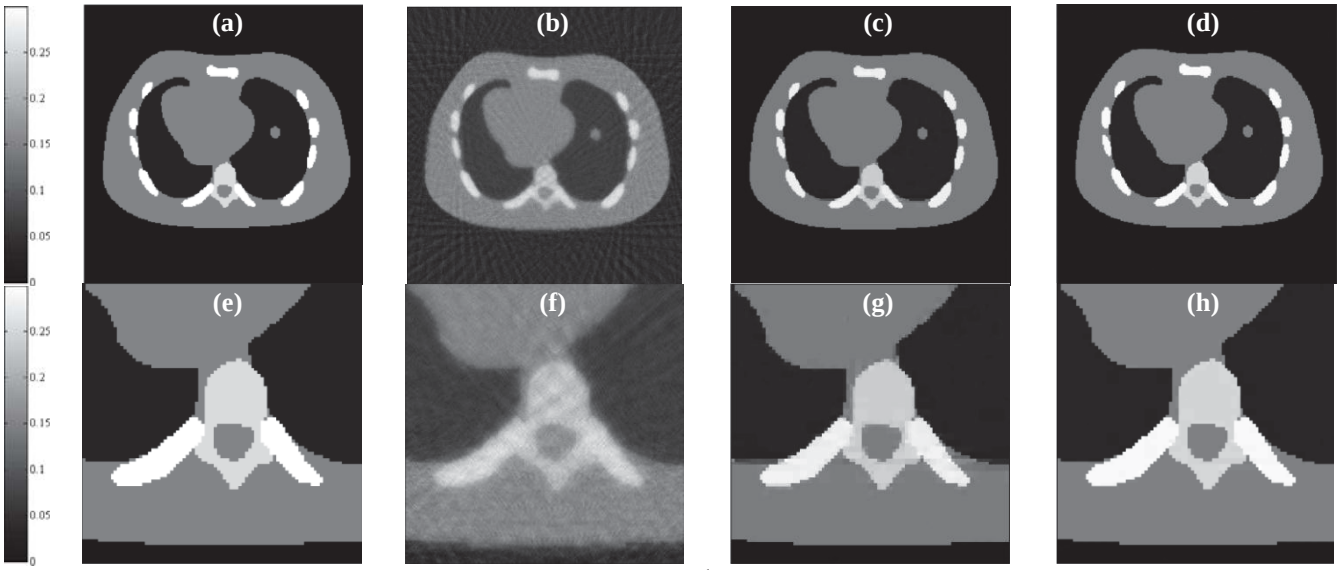


Fig. 2. Reconstructed 512×512 NCAT phantom images (in cm^{-1}) from a 72×724-view noisy sinogram (19.89%): (a), (e) Original image, (b), (f) LS reconstruction ($\mathbf{x}^{(0)}$), (c), (g) l_1-regularized reconstruction, and (d), (h) $l_{\text{p},\varepsilon}$-regularized reconstruction.

Neumann-boundary condition can be more appropriate [16]-[18]. For the same number of iterations, the DCT preconditioner accelerates convergence by obtaining a more accurate solution for the TBT-like matrix, even for a strongly non-symmetric PSF [17], [18].

From [13], we can approximate $\mathbf{A}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{A}$ as follows:

$$\mathbf{A}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{A} \approx \mathbf{G}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{G}, \tag{18}$$

where $\mathbf{G} = diag(\mathbf{g})$, $g_j = \sqrt{\sum_{i=1}^{m} a_{ij}^2 \lambda_i / \sum_{i=1}^{m} a_{ij}^2}$, and $j = 1, \ldots, n$. Applying (18) yields the following approximation of $\mathbf{K}$:

$$\mathbf{K} = \mu\mathbf{A}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{A} + \gamma\mathbf{I} \approx \mathbf{G}(\mu\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{G}^{-2})\mathbf{G}. \tag{19}$$

Based on the DCT, (19) becomes

$$\mathbf{K} \approx \mathbf{G}(\mu\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{G}^{-2})\mathbf{G} \approx \mathbf{G}\mathbf{C}^{\mathrm{T}}\mathbf{\Omega}\mathbf{C}\mathbf{G}, \tag{20}$$

where $\mathbf{C}$ is the orthogonal two-dimensional DCT matrix and

$$\mathbf{\Omega} = diag\{dct_2[(\mu\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{G}^{-2})\delta_{(0,0)}]\}. \tag{21}$$

Finally, the dDCT preconditioner becomes

$$\mathbf{M} = \mathbf{G}^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{\Omega}^{-1}\mathbf{C}\mathbf{G}^{-1}. \tag{22}$$

This is expected to be a closer approximation to $\mathbf{K}^{-1}$ than the diagonal, FFT, or dFFT preconditioners [13]. The additional multiplications associated with $\mathbf{G}^{-1}$ are negligible relative to the two-dimensional DCT. Further, while the cost and storage requirements of the DCT are comparable to the FFT, savings can be achieved by use of real arithmetic rather than complex.

III. Simulation Results and Discussions

Reconstruction algorithms were tested on a two-dimensional 512×512 NCAT chest phantom image with isotropic voxel size of 1 mm. The sinogram was generated with a $\angle \times 724$ fan-beam scanner, using $\angle = 60$, 72 or 90 views. The 724 rays encompass the entire image at 0° and are uniformly spaced. The ring diameter is assumed to be 1.024 m and the rotation range is 360°. For simulation of noise, a compound additive Poisson-Gaussian noise is modeled as

$$\rho_i \sim Poisson(\rho_0 \exp(-[\mathbf{A}\mathbf{x}]_i)) + \mathcal{N}(0, \sigma^2), \tag{23}$$

where $\rho_0 = 5 \times 10^6$ (the number of incident photons, chosen relatively high), the electronic noise variance $\sigma^2 = 250$, and $i = 1, \ldots, m$. We assume that 100 keV photons (chosen relatively high) are emitted from the x-ray tube and the linear attenuation coefficient of the bone $\approx 0.3\ \text{cm}^{-1}$ [19].

For the initial guess the least-square (LS) solution was obtained as $\mathbf{x}^{(0)} = inv(\mathbf{A})\mathbf{y}$ by CG with 100 iterations. For l_1 minimization, $\mu = 10^{-4}$, $\gamma = 500$. For $l_{\text{p},\varepsilon}$ minimization, $\mu = 10^{-4}$, $\gamma = 500$, $p = 5 \times 10^{-4}$, and $\varepsilon = 5 \times 10^{-4}$. The parameters μ and γ were experimentally determined based on best reconstruction accuracy for l_1-norm minimization. The number of iterations for PCG is 2 with warm starting. For $\mathbf{\Psi}$, instead of a discrete gradient transform, a discrete Haar orthogonal wavelet transform with filter size of 4 was used, because it is known better for less-clear edges and less-uniform contrast. The stopping criterion can be practically calculated as and the algorithm stops when it meets the target tolerance ϵ^2:

$$tol(\mathrm{k}) = \frac{\left\|\mathbf{A}\mathbf{x}^{(\mathrm{k})} - \mathbf{y}\right\|_{\mathbf{\Lambda}}^2}{\|\mathbf{y}\|_2^2}. \tag{24}$$

Performance was evaluated using $\log_{10}\left(RMSE(\mathbf{x}^{\text{true}}, \mathbf{x}^{(\mathrm{k})})\right)$.

A. Reconstruction with Fewer Measurements

Images were reconstructed from (60 views) 16.57%, (72) 19.89%, and (90) 24.86% of the number of target image pixels, based on l_1- and $l_{\text{p},\varepsilon}$-regularized statistical CS reconstruction. In reconstruction accuracy, the $l_{\text{p},\varepsilon}$-reconstruction outperforms the l_1 approach and accomplishes almost exact image reconstruction with only 19.89% of the measurements (Fig. 2 and Table 1). Fig. 3(b) shows that a larger number of measurements results in faster convergence and smaller error. Note that one cannot expect always to find a global minimum from the $l_{\text{p},\varepsilon}$ algorithm. Use of the l_1 solution as a starting point may produce a more effective solution as shown in Fig. 3(c).

B. Convergence Rate of Reconstruction

Fig. 3(a) shows that the dDCT preconditioner exhibits faster convergence than the dFFT, FFT, and diagonal preconditioners,

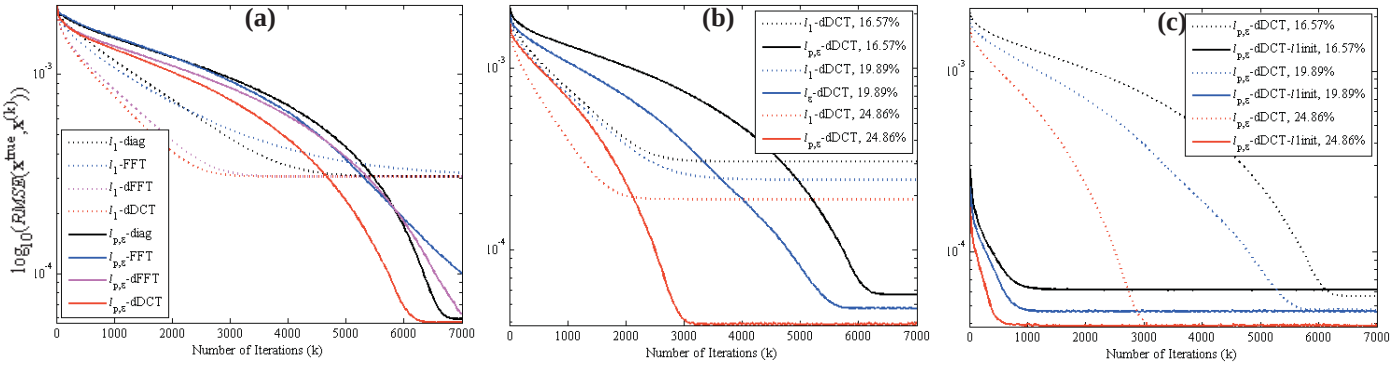


Fig. 3. Convergence of l_1- and $l_{p,\varepsilon}$-regularized statistical reconstructions for NCAT phantom: (a) Comparison of different preconditioning methods (from 60×724-view sinogram, 16.57%). (b) Convergence (with dDCT preconditioning) as function of number of views. (c) Convergence of $l_{p,\varepsilon}$-regularized reconstruction (with dDCT) with $\mathbf{x}^{(0)} = \text{LS}$ and l_1 solution.

for both of the l_1- and $l_{p,\varepsilon}$-regularized reconstructions.

The stability and simplicity of the MM algorithm frequently comes at the price of slow convergence: Fig. 3(a)-(b) illustrates the "waterfall" convergence behavior of the $l_{p,\varepsilon}$-regularized reconstruction. This problem can be resolved by *1)* initiation of reconstruction with l_1-regularization and LS initial guess and a switch to $l_{p,\varepsilon}$-regularization with the initial guess being the l_1 solution at a reasonable tolerance (Fig. 3(c)); *2)* properly decreasing the approximation parameter, ε; or *3)* use of an acceleration approach for expectation-maximization.

Table 1
RMSE COMPARISON WITH DIFFERENT ALGORITHMS

Algorithm (dDCT)	Amount of measurement 16.57%	19.89%	24.86%
LS Recon.	2.2388×10^{-3}	2.0436×10^{-3}	1.8198×10^{-3}
l_1 Recon.	3.0701×10^{-4}	2.4350×10^{-4}	1.8985×10^{-4}
$l_{p,\varepsilon}$ Recon.	5.6576×10^{-5}	4.7367×10^{-5}	3.8424×10^{-5}

IV. CONCLUSIONS

A statistical CS x-ray/CT reconstruction approach has been presented using a non-convex penalty function $l_{p,\varepsilon}(0 < p < 1, 0 < \varepsilon \ll 1)$, based on MM, SB, and a dDCT preconditioner. The method can achieve almost perfect image reconstruction from fewer measurements than with l_1 minimization, with faster convergence than using other known preconditioners. Preliminary simulation results with phantom images support these arguments. Future work seeks to accelerate the convergence rate of the MM-motivated algorithm.

REFERENCES

[1] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, 346(9):589-592, 2008.

[2] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization," *Proceedings of the National Academy of Sciences*, 100(5):2197-2202, 2003.

[3] R. Chartrand, and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, 24(3):035020, 2008.

[4] Y. Shen and S. Li, "Restricted p–isometry property and its application for nonconvex compressive sensing," *Advances in Computational Mathematics*, 37(3):441-452, 2012.

[5] D. Ge, X. Jiang, and Y. Ye, "A note on complexity of Lp minimization," *Mathematical Programming manuscript*, 2010.

[6] H. Yu, and G. Wang, "Compressed sensing based interior tomography," *Physics in medicine and biology*, 54(9):2791, 2009.

[7] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical physics*, 35(2):660, 2008.

[8] J. Choi, M. W. Kim, W. Seong, and J. C. Ye, "Compressed sensing metal artifact removal in dental CT," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, IEEE, 334-337, June 2009.

[9] J. B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," in *Proceedings of SPIE* , 6065:264-273, January 2006.

[10] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ1 minimization," *Journal of Fourier Analysis and Applications*, 14(5), 877-905, 2008.

[11] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM Journal on Imaging Sciences*, 2(2):323-343, 2009.

[12] S. Ramani, and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *Medical Imaging, IEEE Transactions on*, 31(3):677-688, 2012.

[13] S. D. Booth and J. A. Fessler, "Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography," in *Image Processing, 1995. Proceedings., International Conference on*, IEEE, 2:441-444, October 1995.

[14] M. S. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed *transaction* costs," *Annals of Operations Research*, 152(1): 341-365, 2007.

[15] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *American Control Conference, 2003. Proceedings of the 2003*, IEEE, 3:2156-2162, June 2003.

[16] P. C. Hansen, J. G. Nagy, and D. P. O'leary, "*Deblurring images: matrices, spectra, and filtering*," vol. 3, Siam, 2006.

[17] M. K. Ng, R. H. Chan, T. F. Chan, and A. M. Yip, "Cosine transform preconditioners for high resolution image reconstruction," *Linear Algebra and its Applications*, 316(1):89-104, 2000.

[18] R. Chan, T. Chan, M. Ng, W. C. Tang, and C. K. Wong, "Preconditioned iterative methods for high-resolution image reconstruction with multisensors," In *Proceedings to the SPIE symposium on advanced signal processing: Algorithms, architectures, and implementations*, 3461: 348-357, July 1998.

[19] J. H. Hubbell and S. M. Seltzer, *Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z= 1 to 92 and 48 additional substances of dosimetric interest* (No. PB--95-220539/XAB; NISTIR--5632). National Inst. of Standards and Technology-PL, Gaithersburg, MD (United States). Ionizing Radiation Div, 1995.

[20] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure." *Applied and Computational Harmonic Analysis*, 22(3): 335-355, 2007.

[21] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓq-minimization for 0<q≤1." *Applied and Computational Harmonic Analysis*, 26(3): 395-407, 2009.

[22] B. K. Natarajan, "Sparse approximate solutions to linear systems." *SIAM journal on computing*, 24(2): 227-234, 1995.

[23] I.Y. Chun and T. Talavage, "Fast Non-Convex Statistical Compressed Sensing MRI Reconstruction Based on Approximated Lp(0<p<1)-Quasi-Norm with Fewer Measurements than Using L1-Norm," *21st Annual Meeting of the International Society for Magnetic Resonance in Imaging*, 2013

C-ARM CT IMAGE RECONSTRUCTION FROM SPARSE PROJECTIONS

J. Xu and B. M. W. Tsui

I. INTRODUCTION AND MOTIVATION

C-arm CT has seen increased utilization in the interventional suite and the operation room, due to its flexible gantry design and 3D real-time navigation and visualization capability [1]. In many studies, C-arm CT must be performed multiple times on the same patient before, during, and after treatment. In longitudinal studies, a patient or an animal needs to be imaged over a period of time to follow the pathological and physiological changes. In these situations, the increased radiation dose to the imaged subject has become a concern. One method to reduce the radiation dose is to reduce the number of angular projections (sparse projection). The conventional FDK agorithm is unable to handle the sparse projection well; the reconstructed images suffer from severe streaking artifacts. Iterative image reconstruction algorithms [2], [3] that use prior information of the object have been proposed to reduce the image artifacts and maintain the same or similar image quality.

Inspired by the edge-preserving and the non-local mean filters in image denoising, recently non-local and patch-based prior functions have been proposed for emission [4] and x-ray CT image reconstruction [5]. It was found in [4] that the patch-based prior functions are less sensitive to the hyperparameter values. And prior functins that use non-local neighborhood can better utilize the similarity among different regions within the reconstruction volume.

We propose a penalized image reconstruction method using non-local prior functions for sparse view image reconstruction in C-arm CT. The non-local neighborhood is defined from a prior image. The reconstruction algorithm is implemented on the graphical processing unit (GPU) and evaluated using cardiac C-arm animal CT scans. We investigate modifications to well-known optimization algorithms that handle the non-negativity constraint and improve the convergence speed.

II. METHOD

A. Problem formulation

The objective function is formulated as follows.

$$\Phi(x) = \frac{1}{2}\|y - Hx\|^2 + \beta U(x) \quad (1)$$

$$\text{where} \quad U(x) = \sum_{i,j \in N_i} w_{ij}(\tilde{x})\psi(x_i - x_j) \quad (2)$$

$$x \geq 0$$

In eqn. (1), x is the reconstructed image, H is the system matrix. The first term on the right hand side of eqn. (1) is the squared difference between the measured y and the estimated projection data Hx. The prior image $\tilde{x}$ is used to define the similarity weight $w_{ij}(\tilde{x})$ between voxels i and j in the reconstructed image x. We define the weights using local patches around voxel i and j in the prior image [6].

$$w_{ij}(\tilde{x}) = \exp\left\{-\frac{\|\tilde{x}_{[i]} - \tilde{x}_{[j]}\|^2}{\delta^2}\right\} \quad (3)$$

The notation $[i]$ and $[j]$ denote a local patch centered on voxels i and j, and $\|\cdot\|$ is the usual Euclidean norm of a vector.

Assuming the prior image $\tilde{x}$ and the reconstructed image x have similar neighborhood structures, the voxel-dependent weights $w_{ij}(\tilde{x})$ then enable the use of a nonlocal neighborhood N_i in eqn. (2). For the ith voxel, $w_{ij}(\tilde{x})$ assigns larger weights to the jth voxels if j and i have similar patch values in the prior image, though j can be spatially distant from i.

We use the quadratic prior function $\psi(r) = r^2$ in this work. The conventional, local neighborhood quadratic priors tend to over-smooth the reconstructed image since the weight assignments only consider spatial distance between neighboring voxels. On the other hand, the quadratic prior in eqn. (2) uses distant neighborhoods that are determined by the patch similaries in the prior image.

The constants $\beta \geq 0$ in eqn. (1) and $\delta > 0$ in eqn. (3) are design parameters, The parameter δ controls the sensitivity of $w_{ij}(\tilde{x})$ with respect to the voxel difference in the prior image $\tilde{x}$. The smaller the δ, the more sensitive $w_{ij}(\tilde{x})$. The parameter β controls the overall balance between data agreement and the smoothness constraint.

Unlike clinical CT scanners that can achieve sub-second gantry rotation, C-arm systems have a much slower gantry speed, 4-5 secs for a half-scan acquisition. In cardiac applications, in order to reduce motion artifacts, four ECG-triggered alternating direction half-scans (sweeps) are acquired from which projection data at a selected cardiac phase are used for image reconstruction [7]. The averaged image reconstructed from the multiple half-scans has motion artifacts, but the neighborhood structures is similar to that of the selected phase. We use as the prior image $\tilde{x}$ the average of the FDK reconstructed images of all four sweeps. More precisely,

$$\tilde{x} = \frac{1}{4}\sum_{i}^{4} \text{FDK}_i([y_i]).$$

Here i is the scan index, and $[y_i]$ is the projection data of the ith scan.

B. The reconstruction algorithms

Our objective function is in the form of least squares minimization with a quadratic but non-local penalization term. A generic reconstruction algorithm can be written as

$$x_{k+1} = x_k + \alpha d_k, \quad (4)$$

where x_k is the reconstruction at iteration k, d_k is the descent direction for the objective function eqn. (1) at iteration k, and $\alpha > 0$ the step size.

Jingyan Xu and Benjamin M. W. Tsui are with the Division of Medical Imaging Physics, Department of Radiology, Johns Hopkins University. E-mail: {jxu,btsui1}@jhmi.edu.

1) SSF: One special form of eqn. (4) is the following:

$$x_{j,k+1} = \lfloor x_{j,k} - \frac{1}{c + 8\beta \sum_{j \in N_i} w_{ij}(\tilde{x})} \nabla\Phi(x_k) \rfloor_+, \quad (5)$$

It can be derived using the separable surrogate function (SSF) approach. If the constant c satisfies that $cI - H^T H \succ 0$, then the objective function in eqn. (1) decreases monotonically. The minimal value of c is $\lambda_{\max}(H^T H)$, the largest eigenvalue of the matrix $H^T H$, and c can be estimated using the power iteration [8]. In each iteration, there is one forward and one back projection to evaluate the gradient $\nabla\Phi(x_k)$.

2) PCD1: To improve the convergence speed, we consider

$$x_{k+1} \quad = \quad x_k - \alpha M^{-1} \nabla\Phi(x_k), \quad (6)$$

where $M \succ 0$ is a symmetric, positive-definite preconditioner matrix. If we let

$$M = \mathrm{diag}(H^T H) + 8\beta\, \mathrm{diag}(\sum_{j \in N_i} w_{ij}(\tilde{x})), \quad (7)$$

then the descent direction $d_k = -M^{-1}\nabla\Phi(x_k)$ coincides with the parallel coordinate descent (PCD) direction [9]. In other words, each element of the vector d_k is the optimal coordinate increment had a one-coordinate-a-time updating scheme enforced (as in coordinate descent) at the current iteration x_k. To enforce the non-negativity constraint $x_k \in R_+$ for all k, the iteration in eqn. (6) is modified as follows.

$$d_k \quad \leftarrow \quad \max\left\{-M^{-1}\nabla\Phi(x_k), -x_k\right\} \quad (8)$$

$$x_{k+1} \quad = \quad x_k + \alpha d_k, \ 0 \le \alpha \le 1 \quad (9)$$

The $\max$ in eqn. (8) is taken component-wise. With such a modification, each component of d_k is still a descent direction. Since both $x_k \in R_+$ and $x_k + d_k \in R_+$, we have $x_k + \alpha d_k \in R_+$ for $0 \le \alpha \le 1$.

To calculate the step size α, we need to calculate Hd_k. This result can be used to directly obtain $Hx_{k+1} = Hx_k + \alpha H d_k$, which is needed in $\nabla\Phi(x_{k+1})$. There is still one forward and back projection per iteration in eqn. (9).

3) PCD2: To further accelerate the convergence speed, we consider a special case of the subspace method [9], [10] that extends the descent space from 1-dimensional to 2-dimensional. Using the same preconditioner M in eqn. (7) and d_k in eqn. (8), we have

$$p_k \quad = \quad x_k - x_{k-1} \quad (10)$$

$$x_{k+1} \quad = \quad x_k + \alpha d_k + \beta p_k. \quad (11)$$

Compared with eqn. (6), eqn. (11) retains the previous change direction p_k. If we do not consider the non-negativity constraint, the algorithm in eqns. (10)-(11) is equivalent to the (preconditioned) conjugate gradient algorithm [11]. To incorporate the non-negativity constraint, we consider two cases. (1) If $x_k + p_k \in R_+$, then we limit the step sizes such that

$$0 \le \alpha \le 1, \ 0 \le \beta \le 1, \ \alpha + \beta \le 1. \quad (12)$$

In this case, since $x_k \in R_+$, $x_k + d_k \in R_+$, and $x_k + p_k \in R_+$, we have $x_k + \alpha d_k + \beta p_k \in R_+$. (2) If $x_k + p_k \notin R_+$, we let $\beta = 0$ and the subspace method reduces to eqn. (6). Similar to eqn. (9), calculating the step size β does not involve forward or back projection, since the needed quantity $Hp_k = Hx_k - Hx_{k-1}$ is readily available from the previous iteration. There is still one forward and back projection operation in eqn. (10).

Both the preconditioned gradient method and the subspace method require $\mathrm{diag}(H^T H)$. We derive an analytic approximation of $\mathrm{diag}(H^T H)$ in App. A.

III. Reconstruction studies

We applied the nonlocal quadratic penalized least squares reconstruction method to cardiac C-arm animal CT data. The projection data were acquired using the standard cardiac C-arm CT protocols. The x-ray source to detector distance is 1200 mm, and to the rotation center 750 mm. The flat panel detector has 616x480 matrix size and $(0.616\text{mm})^2$ pixel size. One acquisition consisted of 4 ECG-triggered forward and backward half-scans. Each half-scan had 235 projection views at 0.8° increment. From the four acquisitions, one projection dataset at a selected cardiac phase, of 235 views at 0.8° increment, is also available. We subsampled this dataset by 5 and reconstructed both the full and the reduced datasets using the FDK and the proposed method.

All image reconstructions were performed with 256^3 matrix size and 1mm^3 voxel size. For the iterative reconstruction method, we varied the penalization weight β and the prior sensivitivity parameter δ. The forward and backprojection operations, and the neighborhood weights $w_{ij}(\tilde{x})$ were calculated on the GPU. We used a patch size of 3 x 3 x 3 to determine the local neighborhood similarity [cf. eqn. (3)]. Ideally, the nonlocal region [N_i in eqn. (2)] should encompass the entire image. But the computation time of the prior weights $w_{ij}(\tilde{x})$ will quickly exceed that of the forward and backprojection. This is now the limiting factor of our nonlocal region size. Currently we used a non-local neighborhood N_i consisting of 3 1-dimensional array of voxels, with 11 voxels per orthogonal axis centered on voxel i. The reconstruction time using the above parameters is around 40 sec per itertation.

We plot in Fig. 1 the objective function value as a function of iteration numbers of the three methods in Sec. II-B. The reconstruction results using PCD2 are shown in Fig. 2. It can be seen that PCD2 has the fastest convergence among the three methods, especially at early iterations. At later iterations ($\sim$100), the difference between PCD1 and PCD2 is negligible, and both are significantly faster than the separable surrogate function approach (SSF). The sample reconstruction results demonstrate that the proposed least squares method with the nonlocal quadratic prior function can preserve the sharp image features and effectively reduce the streaking artifacts from sparse projection.

IV. Summary

We developed a nonlocal quadratic penalized least squares method for image reconstruction from sparse projections. Among the three optimization algorithms that we investigated, the modified (with non-negativity constraint) subspace method (PCD2) has the fastest convergence speed, compared with the preconditioned gradient method (PCD1) and the separable surrogate function (SSF) approach. All three variants take

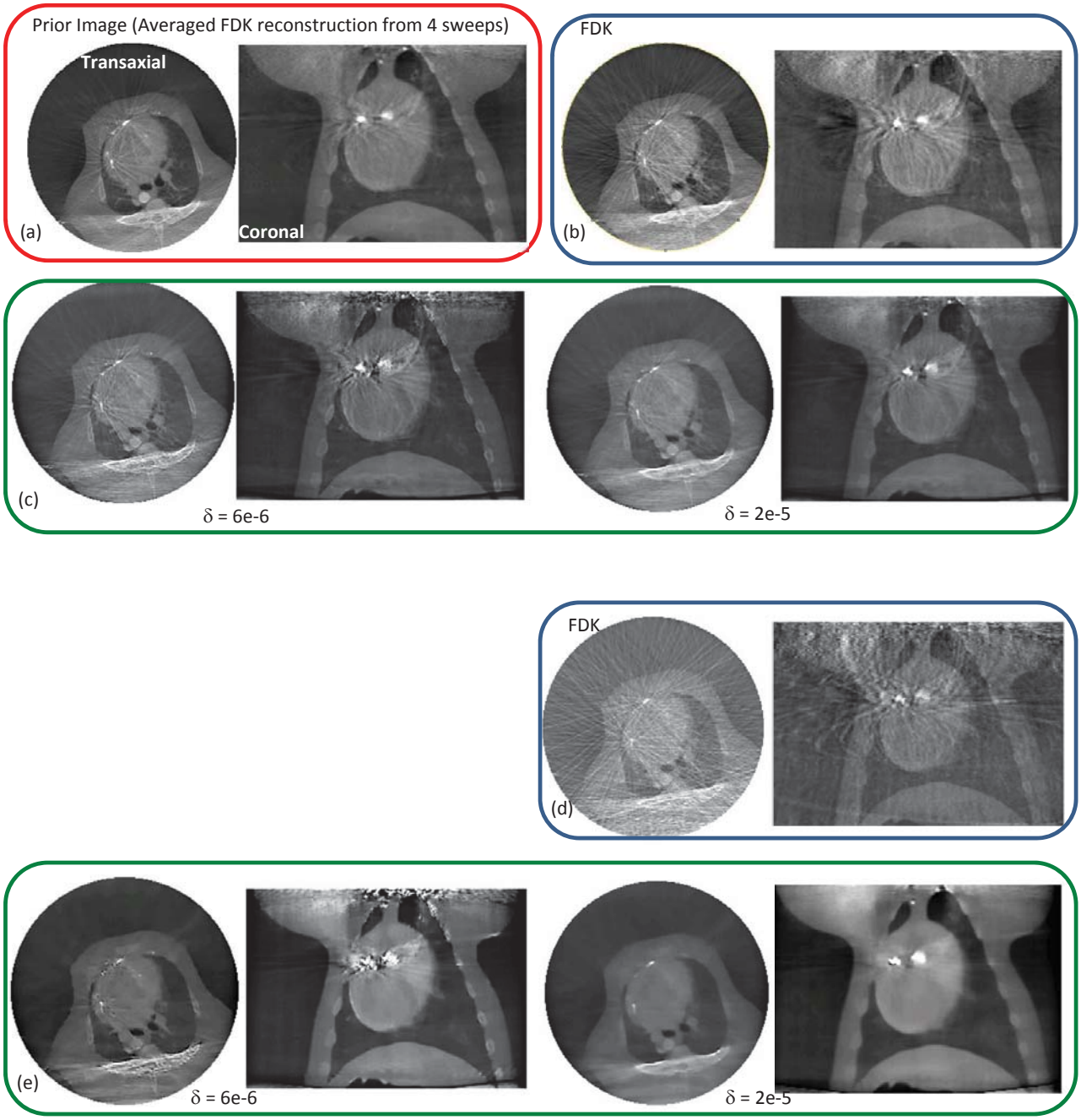


Fig. 2: (a) The prior image was generated using the averaged reconstruction from all 4 sweeps. Reconstruction results from the in-phase 235 projection views using (b) FDK algorithm and (c) the proposed method at two δ values. (d) and (e) are parallel to (b) and (c) but use 47 projection views that were subsampled by 5 from the 235 views. All iterative reconstruction results are at iteration 500, and with $\beta = 500$ and two values of δ.

into account the non-negativivity constraints in image updates. We also proposed an analytic approximation of the matrix $\mathrm{diag}(H^T H)$ needed in PCD1 and PCD2.

Appendix A
An analytic expression of the diag{HtH}

To calculate $\mathrm{diag}(H^T H)$, we notice that each of the diagonal elements can be regarded as the total "energy" of the point response function. Using the cone beam projection equation,

$$g(\lambda, u, w) = \int_0^\infty dt f(\vec{a}(\lambda) + t\vec{\theta}), \tag{13}$$

$$\begin{aligned} \vec{\theta} &= \frac{u e_u(\lambda) + w e_w + D e_v(\lambda)}{\sqrt{u^2 + w^2 + D^2}} \\ e_v &= [-\cos\lambda, -\sin\lambda, 0] \\ e_u &= [-\sin\lambda, \cos\lambda, 0], \quad e_w = [0, 0, 1] \\ \vec{a}(\lambda) &= [R\cos\lambda, R\sin\lambda, 0] \end{aligned} \tag{14}$$

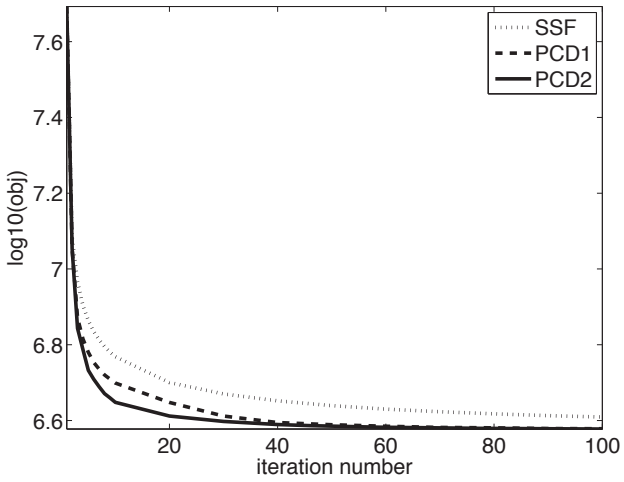


Fig. 1: The objective function as a function of the iteration numbers of the three reconstruction algorithms in Sec. II-B. Both PCD1 and PCD2 outperform the separable surrogate function (SSF) approach.

we can calculate the "total energy" of a point object located at $\boldsymbol{x}_0$, $f_0(\boldsymbol{x}) = \delta(\boldsymbol{x} - \boldsymbol{x}_0)$, in the object space. Let

$$h_\lambda \stackrel{\triangle}{=} \int du\ dw\ g_0(\lambda, u, w) = \int du\ dw\ dt\ f_0(\vec{a}(\lambda) + t\vec{\theta}) \tag{15}$$

Using a change of variables, the Jacobian (which is equal to h_λ) from $du\,dw\,dt$ to the Cartesian coordinate system $d\boldsymbol{x}$ can be shown to be

$$h_\lambda = \frac{D^2 \|\boldsymbol{x}_0 - \vec{a}(\lambda)\|}{|[\boldsymbol{x}_0 - \vec{a}(\lambda)] \cdot \vec{e}_v(\lambda)|^3}.$$

Then the needed diagonal element

$$[H^T H]_{i,i} \quad = \sum_{\lambda, s.t. i \text{ visible}} h_\lambda^2.$$

At each voxel location i, the summation is over all the source positions such that the voxel i is visible (projected to the detector surface at that source position). Fig. 3 is a transaxial slice of the $\mathrm{diag}(H^T H)$ matrix used in the reconstruction algorithms PCD1 and PCD2.

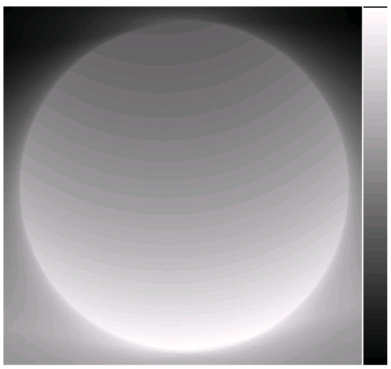


Fig. 3: A transaxial slice of the matrix $\mathrm{diag}(H^T H)$. During the half-scan acquisition, the x-ray source rotates closer to the bottom half (brighter) of the image.

REFERENCES

[1] W. Kalender and Y. Kyriakou, "Flat-detector computed tomography (FD-CT)," *European Radiology*, vol. 17, no. 11, pp. 2767–2779, 2007.

[2] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical physics*, vol. 35, pp. 660–663, Feb. 2008.

[3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, Sept. 2008.

[4] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 2194 –2204, Dec. 2012.

[5] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, "Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior," *Physics in Medicine and Biology*, vol. 57, pp. 7519–7542, Nov. 2012.

[6] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, pp. 490–530, Jan. 2005.

[7] G. Lauritsch, J. Boese, L. Wigstrom, H. Kemeth, and R. Fahrig, "Towards cardiac C-arm computed tomography," *Medical Imaging, IEEE Transactions on*, vol. 25, pp. 922 –934, july 2006.

[8] G. H. Golub and C. F. v. V. Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences*. The Johns Hopkins University Press, 3rd ed., Oct. 1996.

[9] M. Elad, B. Matalon, and M. Zibulevsky, "Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization," *Applied and Computational Harmonic Analysis*, vol. 23, pp. 346–367, Nov. 2007.

[10] A. Miele and J. W. Cantrell, "Study on a memory gradient method for the minimization of functions," *Journal of Optimization Theory and Applications*, vol. 3, pp. 459–470, Nov. 1969.

[11] J. W. Cantrell, "Relation between the memory gradient method and the Fletcher-Reeves method," *Journal of Optimization Theory and Applications*, vol. 4, pp. 67–71, July 1969.

ACKNOWLEDGMENT

The authors would like to thank Drs. Dara Kraitchman and Clifford Weiss for the C-Arm CT data.

Nonlinear diffusion constraints for reconstructing subsampled rotational angiography data

Hélène Langet[1,2,3], Aymeric Reshef[1], Cyril Riddell[1], Yves Trousset[1], Arthur Tenenhaus[3], Elisabeth Lahalle[3], Gilles Fleury[3], and Nikos Paragios[2,4,5]

***Abstract*—Interventional imaging with cone-beam C-arm CT often lacks sufficient sampling. Compressed sensing based reconstruction algorithms have shown promising results to improve image quality in this context using sparsity constraints. Compressed sensing theory by itself assumes random measurements and ℓ_1-penalties. In practice, benefits are seen with uniform subsampling patterns. Here, we investigate substituting ℓ_1 total variation with a nonlinear diffusion constraint and show on a clinical data set that image quality is also improved. This result adds flexibility to the design of CS-based algorithms as C-arm CT images may not be so well approximated by piecewise constant functions.**

***Index Terms*—Rotational angiography, iterative reconstruction, compressed sensing, total variation, nonlinear diffusion**

I. Introduction

Tomographic reconstruction is computed analytically through filtered backprojection (FBP) whose discretization defines sampling requirements. For interventional imaging with C-arm systems, because of the low framerate of the detector, the sampling is not as favorable as what is achieved for diagnostic imaging with CT scanners. Image quality improvement is however much needed to increase the use of tomography in the interventional practice.

As an alternative to FBP, iterative approaches translate the tomographic problem into a discrete problem (see Sec. II-A) to estimate a density mapping that fits the projection data by optimizing a cost function such as weighted least square (WLS). Because the solution cannot be uniquely determined by the data, WLS has to be constrained. The recent developments of the compressed sensing (CS) theory highlighted, in particular, the importance of ℓ_1 penalties for handling subsampled data [1]. These penalties can be combined to WLS through proximal algorithms as shall be recalled in Sec. II-B.

A CS-based algorithm was proposed for reconstructing 'sparse vessels over a non-sparse background' [2]. It solves a series of WLS problems penalized by the ℓ_1-norm of the image combined with positivity, whith decreasing levels of penalization. It was shown to segment the high-intensity vessels and thus significantly mitigate the associated subsampling artifacts, while the background was progressively reintroduced. Such an approach belongs to 'homotopy' strategies whose principle is described in Sec. II-C.

The authors are with 1/ GE Healthcare (Buc, France), 2/ École Centrale Paris, Center for Visual Computing (Châtenay-Malabry, France), 3/ Supélec (Gif-sur-Yvette, France), 4/ École des Ponts ParisTech, Center for Visual Computing (Champs-sur-Marne, France), and 5/ INRIA Saclay, GALEN team (Orsay, France). Corresponding author: cyril.riddell@ge.com. This work was supported by the CIFRE grant n° 936/2009 from the French Association Nationale de la Recherche et de la Technologie (ANRT).

Assuming an image is piecewise constant, many works have exploited penalization by the total variation (TV) for which there exists efficient minimization algorithms. However, C-arm CT images may not be so well approximated by piecewise constant functions. On the other hand, nonlinear anisotropic diffusion (NLAD) presents edge preserving properties similar to TV without relying on the piecewise constant assumption. Section II-D introduces a unified framework through the alternating direction method of multipliers (ADMM) to compute the proximal operators of TV and NLAD. Results on clinical data are presented in Sec. III.

II. Method

A. Discrete tomographic problem

Let us denote $f \in \mathbb{R}^K$ the vector that is associated with the imaged object, where K is the number of voxels in the 3D space. Coefficient $(f)_k$ represents the linear attenuation at voxel k. Let us denote $p \in \mathbb{R}^{J \times N}$ the vector that refers to the set of measurements, where N is the number of angular positions and J is the number of measurements at each incidence. Let us then denote $R \in \mathbb{R}^{J \times N} \times \mathbb{R}^K$ the projection matrix that models the rotational cone beam acquisition of p of f, where a column of R refers to a given voxel, while a row refers to a given measurement. The tomographic reconstruction problem is described by the system of linear equations

$$Rf = p. \tag{1}$$

This work aims at solving the reconstruction problem in the underdetermined case where the measurements are severely subsampled (N small), so that $J \times N \ll K$. Hence, there exists an infinity of solutions that are compatible with Eq. 1.

B. Penalized weighted least square (PWLS)

The selection of clinically relevant solutions can be carried out through PWLS that consists in minimizing a functional that is the combination of:

- fidelity term $Q_W(f) = \frac{1}{2}(Rf - p)^T W (Rf - p)$, where $.^T$ refers to the transpose of a matrix and W to a positive-definite weighting matrix (e.g. a statistical data noise model);
- and penalty $\chi(Af)$, where χ is a convex function (not necessarily differentiable) and A is a linear operator (e.g. the identity operator as in [2] or a wavelet transform or the gradient).

Let us consider the λ-indexed PWLS problem:

$$(f_{\chi\circ A})_\lambda = \underset{f\in\mathbb{R}^K}{\arg\min}\Big\{Q_W(f)+\lambda\chi(Af)\Big\}, \tag{2}$$

where hyperparameter λ denotes the weight assigned to the penalization. To solve Eq. 2, we rely on proximal splitting as proposed by Combettes et al. [3] and split the optimization process into an explicit gradient step for minimizing Q_W and an implicit step applying the constraint χ on domain A through the proximal operator

$$\operatorname{prox}_{\lambda\chi\circ A}(f_0) \equiv \underset{f\in\mathbb{R}^K}{\arg\min}\left\{\lambda\chi(Af)+\frac{1}{2}\|f-f_0\|_2^2\right\}. \tag{3}$$

The computation of iterate $f^{(i+1)}$ is then given by

$$f^{(i+1)} = \underbrace{\operatorname{prox}_{\tau\lambda\chi\circ A}}_{\text{implicit step}}\big(\underbrace{f^{(i)}-\tau\nabla Q_W(f^{(i)})}_{\text{explicit gradient step}}\big). \tag{4}$$

C. Homotopy

When using ℓ_1-norms, the regularization path is defined as the family of solutions $(f_{\chi\circ A})_\lambda$ of Eq. 2 when varying hyperparameter λ over $[0,+\infty[$. Efron et al. [4] have shown that there exists a series of values (λ_s) with $0=\lambda_1<\cdots<\lambda_{\max}$, such that the regularization path is a piecewise linear function of λ:

$$(f_{\chi\circ A})_\lambda = \frac{\lambda-\lambda_s}{\lambda_{s+1}-\lambda_s}(f_{\chi\circ A})_{\lambda_{s+1}} + \frac{\lambda_{s+1}-\lambda}{\lambda_{s+1}-\lambda_s}(f_{\chi\circ A})_{\lambda_s},$$
$$\text{where } \lambda_{s+1}\geq\lambda\geq\lambda_s.$$

Given that the higher the constraint, the faster the convergence, homotopy strategies were developed to compute $(f_{\chi\circ A})_{\lambda_S}$ by computing a sequence of S solutions $((f_{\chi\circ A})_{\lambda_s})$, where λ_s varies from large to small values such that $\lambda_1\geq\lambda_s>\lambda_{s+1}\geq\lambda_S$ and $(f_{\chi\circ A})_{\lambda_s}$ is used as initialization to computing $(f_{\chi\circ A})_{\lambda_{s+1}}$. In practice, each intermediate solution is only approximated. The number of stages S is not known and is instead set as a parameter that shall reflect some *a priori* knowledge on the image structure.

The homotopy approach is relevant to low angular sampling because the minimization of an underdetermined LS criterion strongly depends on the initialization. In particular, if initialized by the true image, sampling artifacts are not reintroduced by the data fidelity term. Thus, minimization at a given λ identifies a sparse approximation of the solution that best fits the data with a level of sparsity that is proportional to λ. Any sparse structure that actually belongs to the solution will be kept when applying successive lower λ-values so that the final solution verifies the data fidelity term. On the contrary, if the approximation does not fit the data, it will be removed by the subsequent minimizations. This is why this approach is not relevant to fully sampled noisy data because then the minimization of the LS criterion does not depend on initialization. In this case, knowledge of the solution does not eliminate the noise of the LS minimum so that the goal of regularization is indeed to reach a biased solution to avoid an overfitting of the data fidelity term.

We shall thus not use matrix W to model the noise in the data, as it would slow the convergence for little benefit, at least for strong λ.

D. Computing the proximal operator with ADMM

The proximal formalism covers all standard image processing operations that are solution of a LS-based variational approach. When constraint χ is the ℓ_1-norm, and when $A=I$, the proximal operator can be expressed as a soft thresholding shrinkage operator:

$$\operatorname{prox}_{\lambda\|.\|_1}(f) = \operatorname{shrink}_{\lambda\|.\|_1}(f) \tag{5}$$

where

$$\big(\operatorname{shrink}_{\lambda\|.\|_1}(f)\big)_k = \frac{f_k}{|f_k|}\max(|f_k|-\lambda,0).$$

If we denote $g=(g^x,g^y,g^z)$ the gradient of image f, its ℓ_1-norm can be defined either as anisotropic with

$$\|g\|_1 = \sum_{k=0}^{K}\|g_k\|_1 = \sum_{k=0}^{K}|g_k^x|+|g_k^y|+|g_k^z|$$

or isotropic with

$$\|g\|_{1,2} = \sum_{k=0}^{K}\|g_k\|_2 = \sum_{k=0}^{K}\sqrt{|g_k^x|^2+|g_k^y|^2+|g_k^z|^2}$$

In the case of the isotropic norm, we have

$$\operatorname{prox}_{\lambda\|.\|_{1,2}}(f) = \operatorname{shrink}_{\lambda\|.\|_{1,2}}(f), \tag{6}$$

using the generalized shrinkage formula

$$\big(\operatorname{shrink}_{\lambda\|.\|_{1,2}}(g)\big)_k = \frac{g_k}{\|g_k\|_2}\max(\|g_k\|_2-\lambda,0).$$

When A is orthogonal and of unit norm (e.g. wavelet transform), one easily proves that

$$\operatorname{prox}_{\lambda\chi\circ A}(f) = A^{-1}\operatorname{prox}_{\lambda\chi}(Af).$$

To enforce the piecewise constant nature of an image (i.e. to preserve the edges and smooth out other areas), a common approach is to penalize the image gradient with $\chi=\|.\|_1$ or $\chi=\|.\|_{1,2}$, using $g=Af=\nabla f$. Goldstein and Osher [5] suggested the Alternating Direction Method of Multipliers (ADMM) to calculate $\operatorname{prox}_{\lambda\chi\circ A}(.)$ even though A is not invertible, by iteratively solving a quadratic problem involving A and replacing $\operatorname{prox}_{\lambda\chi\circ A}(.)$ by $\operatorname{prox}_{\lambda\chi}(.)$, which in this case are the above defined shrinkage operators.

ADMM consists in splitting the problem into a joint minimization over both the image itself f and variable $v=Af$, thus leading to an optimization problem of the form

$$\operatorname{prox}_{\lambda\chi\circ A}(f_0) = f^* \tag{7}$$

where f^* is computed according to

$$(f^*,v^*) = \underset{(f,v)}{\arg\min}\Big\{\lambda\chi(v)+\frac{\mu}{2}\|v-Af\|_2^2 + \frac{1}{2}\|f-f_0\|_2^2\Big\} \tag{8}$$

with $\mu>0$. A generalized inversion of A is accomplished through term $\|v-Af\|_2^2$. ADMM alternatively minimizes (8) along direction f with v fixed, then along direction v with f fixed. This requires to add a variable b that will track the inversion error $v-Af$.

When v and b are fixed to $v^{(j)}$ and $b^{(j)}$ the solution

$$f^{(j+1)} = \arg\min_f \left\{ \frac{\mu}{2}\|v^{(j)} - (Af + b^{(j)})\|_2^2 + \frac{1}{2}\|f - f_0\|_2^2 \right\} \quad (9)$$

is quadratic and its implementation is *independent* from both the regularization term λ and the regularization penalty χ. When f and b are fixed to $f^{(j+1)}$ and $b^{(j)}$, the computation of $v^{(j+1)}$ is made via

$$v^{(j+1)} = \text{prox}_{\frac{\lambda}{\mu}\chi}(Af^{(j+1)} + b^{(j)}). \quad (10)$$

Vector b is then simply updated as:

$$b^{(j+1)} = b^{(j)} + (Af^{(j+1)} - v^{(j+1)}). \quad (11)$$

In the following we implement three filters with ADMM: the anisotropic TV, the isotropic TV and the anisotropic diffusion.

1) Anisotropic Total Variation (TV-1): The anisotropic TV filter was pioneered by Rudin et al. [6] and is the proximal operator that minimizes the anisotropic ℓ_1-norm:

$$\text{prox}_{\lambda\text{TV1}}(f_0) = \arg\min_f \left\{ \lambda\|\nabla f\|_1 + \frac{1}{2}\|f - f_0\|_2^2 \right\}. \quad (12)$$

ADMM is thus applied with $\chi = \|.\|_1$, $A = \nabla$ and $v = g$. Eq. 10 leads to the shrinkage formula in the gradient domain

$$g^{(j+1)} = \text{shrink}_{\frac{\lambda}{\mu}\|.\|_1}(\nabla f^{(j+1)} + b^{(j)}), \quad (13)$$

2) Isotropic Total Variation (TV-2): The isotropic TV filter is the proximal operator that minimizes the isotropic ℓ_1-norm:

$$\text{prox}_{\lambda\text{TV2}}(f_0) = \arg\min_f \left\{ \lambda\|\nabla f\|_{1,2} + \frac{1}{2}\|f - f_0\|_2^2 \right\} \quad (14)$$

and ADMM is now applied with $\chi = \|.\|_{1,2}$, $A = \nabla$ and $v = g$. Eq. 10 now leads to the generalized shrinkage formula in the gradient domain

$$g^{(j+1)} = \text{shrink}_{\frac{\lambda}{\mu}\|.\|_{1,2}}(\nabla f^{(j+1)} + b^{(j)}). \quad (15)$$

Note that with TV, parameter λ plays both the role of a threshold parameter in the shrinkage operators and a penalization weight in the minimization problem. Thus, it both defines an image scale (only edges whose gradients are greater than λ remain after the shrink) and an amount of filtering (the higher λ, the flatter the image).

3) Nonlinear anisotropic diffusion (NLAD): TV-2 minimization is equivalent to solving

$$\arg\min_f \left\{ \frac{\lambda}{4}\|\varphi(\|\nabla f\|_2^2)\|_2^2 + \frac{1}{2}\|f - f_0\|_2^2 \right\}, \quad (16)$$

where $\varphi(\delta) = 2\sqrt[4]{\delta}$, and $c(\delta) = \varphi(\delta)\varphi'(\delta) = \frac{1}{\sqrt{\delta}}$ can be seen as some degenerated diffusivity map with an infinite diffusion in regions where $\|\nabla f\|_2^2$ tends to zero.

In the case of NLAD, function φ and c are replaced by φ_γ and c_γ such that c_γ satisfies conditions discussed in [7]. Scalar γ plays the role of a scale parameter [8]: all gradients whose norms are greater than γ are considered as edges and are preserved by the map, while λ tells the strength of the diffusion between the edges. Hence, the amount of edges we want to keep is *decorrelated* from the amount of filtering between the edges.

NLAD is again a proximal operator defined by

$$\text{prox}_{(\lambda,\gamma)\text{NLAD}}(f_0) = \arg\min_f \left\{ \frac{\lambda}{4}\|\varphi_\gamma(\|\nabla f\|_2^2)\|_2^2 + \frac{1}{2}\|f - f_0\|_2^2 \right\}. \quad (17)$$

Interestingly, ADMM can again be used here with $A = \nabla$, $v = g$ and $\chi = \|\varphi(\|.\|_2^2)\|^2$. The quadratic problem of Eq. 9 is unchanged, but Eq. 10 now leads to the optimality condition on the image gradient update $g^{(j+1)}$:

$$\left[1 + \frac{\lambda}{\mu}c_\gamma(\|g^{(j+1)}\|_2^2)\right] g^{(j+1)} = \nabla f^{(j+1)} + b^{(j)}. \quad (18)$$

Eq. 18 is nonlinear and hard to solve in general. To simplify the problem we adopt the point of view of Chan and Shen [9]. We keep c_γ as a fixed map from a previous approximation of the image gradient $g^{*(j')}$:

$$c_\gamma^{(j')} = c_\gamma(\|g^{*(j')}\|_2^2). \quad (19)$$

Eq. 18 is now linear and leads to:

$$g^{(j')(j+1)} = \frac{1}{1 + \frac{\lambda}{\mu}c_\gamma^{(j')}}(\nabla f^{(j+1)} + b^{(j)}). \quad (20)$$

Thus, instead of shrinking $\nabla f^{(j+1)} + b^{(j)}$ as in TV, NLAD weights it by a nonuniform multiplicative factor varying from $\frac{1}{1+\lambda/\mu}$ to 1, depending on map c_γ. When the scheme reaches convergence to some image gradient $g^{*(j'+1)}$, the map is updated from the current solution and problem (17) is solved again, but with diffusivity map $c_\gamma^{(j'+1)}$.

Given NLAD parametrization, a regularization path can mimic TV homotopy by varying γ from $\gamma_{\max} > 0$ to 0, while keeping λ fixed at a high value.

III. Experiments and Results

In the following, we consider a clinical non-injected acquisition of 600 projections of the head and provide tomographic reconstructions with homotopy of a uniformly subsampled subset of 35 projections, which is far below a clinically relevant number of projections but exemplifies how efficient the algorithms can be. We rely on a data fidelity term where $W = D$ is the ramp filter that is positive and diagonal in the Fourier domain. We consider a series of $S = 11$ minimization problems, each of them being only approximately solved with a gradient step set to $\tau = 0.9$ and a single iteration in the proximal splitting. We use the same ADMM framework for both TV and NLAD minimizations with $\mu = 1$. The quadratic subproblem of Eq. 9 is solved with 6 Gauss-Seidel iterations. ADMM algorithm ends after 50 iterations or if the condition $\|f^{(j+1)} - f^{(j)}\|/\|f^{(j)}\| < 0.5\%$ is met. Both methods are compared to classical non constrained reconstruction with Feldkamp (FDK) algorithm.

For TV penalties, parameter λ_s decreases linearly from 600 to 0 in steps of 60. For NLAD, we keep a constant diffusion

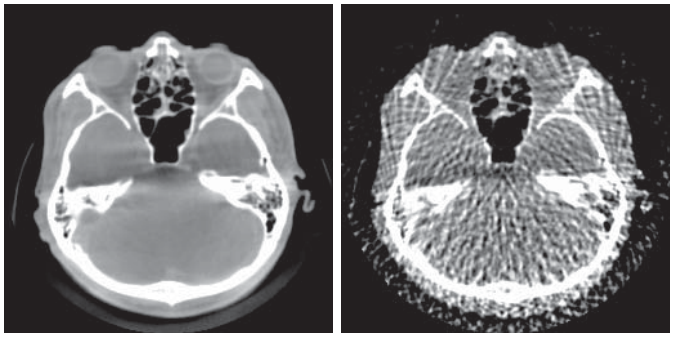


Fig. 1. FDK reconstructions. WW = 500 HU - WL = 500 HU. Left: from 600 projections. Right: from 35 projections.

strength λ equal to 1000 and we decrease diffusivity map threshold γ_s from 100 HU to 0 HU in steps of 10 HU. We use Weickert's diffusivity function [8]

$$c_\gamma(s) = \begin{cases} 1 & (s \leq 0) \\ 1 - \exp\left(\frac{-3.315}{(s/\gamma^2)^4}\right) & (s > 0) \end{cases} . \qquad (21)$$

We compute the first diffusivity map from a smooth version of the input in order to avoid false edge detections. We found that the map converged within 2 passes.

Both TV and NLAD starting parameters are chosen to provide similar filtering strength while removing all streak artifacts.

Figure 1 (left) shows the FDK reconstruction from 600 projections. Due to various limitations in the measurement process, the fully sampled image is not made of pure flat areas as a CT scanner image. Figure 1 (right) shows the FDK reconstruction from 35 projections, while Fig. 2 (top left) shows the LS reconstruction obtained by iterating FDK. Interestingly, iterative FDK and FDK reconstructions differ, meaning that FDK does not fully satisfy the data fidelity term at this low number of projections, with the LS solution being smoother. All nonlinear constraints found solutions to the data fidelity term of much higher image quality. In this example, TV-1 (top right) shows slightly more artifacts than TV-2 (bottom left), while NLAD (bottom right) visually outperformed both TV penalties.

IV. Discussion and conclusion

Compressed sensing based algorithms have been shown to provide clear improvements when reconstructing rotational angiography data that are affected by subsampling. The technical conditions of the theory (ℓ_1-minimization, random measurements) limits its rigorous applicability. Fortunately, streak reduction is obtainable with TV even with a uniform angular subsampling. Here, we replicated a homotopy strategy with proximal operators using an alternative nonlinear LS constraint. For the proposed parametrization on a specific example, the latter constraint outperformed TV, suggesting it better captured the fully sampled solution that is not perfectly piecewise constant. Because the selected constraint is a well-known filter, the procedure is highly intuitive and should allow for an easier design of constraints for subsampled reconstruction. Thanks to ADMM, we were able to change the type of constraints by essentially switching from a gradient shrinkage operation for TV to a nonlinear gradient scaling operation for NLAD. Note, however, that NLAD could have been computed with the scheme proposed by Weickert [8]. This scheme is potentially faster and above all does not require the large amount of memory space necessary for storing ADMM joint variables.

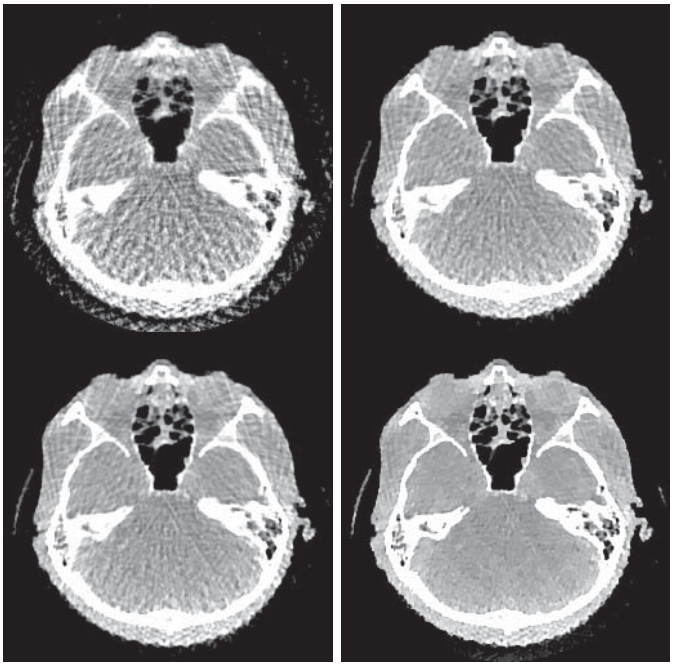


Fig. 2. Iterative reconstructions from 35 projections. WW = 1000 HU - WL = 1000 HU. Top left: LS reconstruction obtained by iterating FDK. Top right: TV-1 reconstruction. Bottom left: TV-2 reconstruction. Bottom right: NLAD reconstruction.

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] H. Langet, C. Riddell, Y. Trousset, A. Tenenhaus, E. Lahalle, G. Fleury, and N. Paragios, "Compressed Sensing Based 3D Tomographic Reconstruction for Rotational Angiography," in *MICCAI*, 2011, vol. 6891, pp. 97–104.

[3] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 2011, vol. 49, ch. 10, pp. 185–212.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

[5] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[7] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, ser. Applied Mathematical Sciences. Heidelberg: Springer, 2006, vol. 147.

[8] J. Weickert, B. M. T. H. Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, vol. 7, pp. 398–410, 1998.

[9] T. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Society for Industrial and Applied Mathematics, 2005.

Robust Motion Estimation for On–Board CBCT Imaging using an Angular Sampling Artifact Model

Marcus Brehm, Pascal Paysan, Markus Oehlhafen, and Marc Kachelrieß

***Abstract*—An additional kV imaging system next to the linear particle accelerator provides valuable information in image–guided radiation therapy (IGRT). However, due to the limited gantry rotation speed during treatment the typical acquisition time is much longer than the patient's breathing cycle resulting in low image quality.**

Motion–compensated image reconstruction is an interesting option and capable of providing high quality respiratory–correlated 4D volumes. The particular challenge is to determine the required motion vector fields for motion compensation.

For reasons of inter–fractional variations and shortest possible duration of treatments, we avoid using knowledge from prior scans and we do not impose specific requirements on the data acquisition. The consequences are that state–of–the–art methods for motion estimation suffer from image artifacts and tend to match artifacts rather than anatomy.

We propose a robust motion estimation method using an angular sampling artifact model that addresses the image artifact problem. The method is a combination of two approaches. One part is a cyclic registration method with temporal constraints like cyclic breathing motion patterns. The second part is a second 4D image series which models the angular sampling artifacts of the gated 4D CBCT but which is free of respiratory motion.

We verified our motion estimation method by motion–compensated reconstructions using simulated rawdata. Furthermore, we successfully processed patient data and the results will be presented at the meeting. A low sensitivity on image artifacts is shown. By using an angular sampling artifact model the robustness is strengthened of the cyclic registration method with temporal constraints. In this way, the motion is accurately estimated and a motion compensation corrects for it.

***Index Terms*—image–guided radiation therapy, cone–beam computed tomography, motion estimation**

I. Introduction

ON–board cone–beam CT (CBCT) imaging provides valuable information in radiation therapy, e.g. for an accurate patient positioning, for recalculation and verification of delivered dose to the patient on the treatment day, and for on–line generation of new treatment plans (c.f. figure 1). However, the maximum gantry rotation speed of 6° per second results in long acquisition times. In thoracic imaging this leads to severe artifacts like motion blurring and streaks in 3D CBCT due to lung and heart motion.

Therefore, gated 4D CBCT has been proposed which consists of a retrospective respiratory binning step followed by independent reconstructions. However, the respiratory correlation results in a large angular spacing between projections used for a single reconstruction. This leads to strong streak artifacts in gated 4D CBCT images.

Marcus Brehm and Prof. Dr. Marc Kachelrieß: Medical Physics in Radiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

Marcus Brehm and Prof. Dr. Marc Kachelrieß: Institute of Medical Physics, Friedrich-Alexander-University (FAU) Erlangen–Nürnberg, 91052 Erlangen, Germany.

Dr. Pascal Paysan and Dr. Markus Oelhafen: Varian Medical Systems, 5405 Baden–Dättwil, Switzerland.

Corresponding author: marcus.brehm@dkfz.de

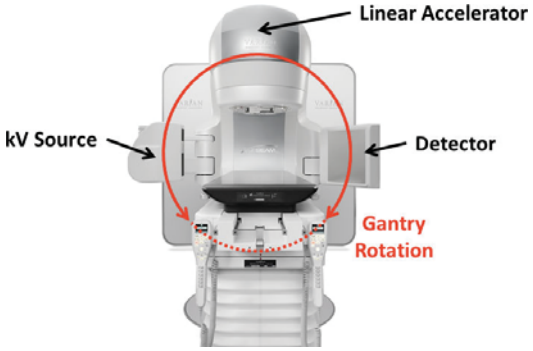


Figure 1. State-of-the-art treatment delivery system with on–board imager for 2D and 3D kV image guidance during radiation therapy.

Several attempts have been made so far to improve image quality of 4D CBCT from on–board scans. Dedicated acquisition techniques with particularly slow, multiple or adaptive gantry rotation seek to reduce the streak artifacts. McKinnon and Bates take a different approach by creating a correction image that shows the same artifacts as the gated 4D CBCT image and a subsequent subtraction of the images significantly reduces those artifacts [1]. Other approaches consider that not all patient regions are moving during the acquisition [2].

Motion–compensated image reconstruction [3] is a promising solution to the streaking problem of respiratory–correlated reconstructions. Here, respiratory motion is compensated via motion vector fields (MVF) and all projection data are thus applied for image reconstruction without a loss in temporal resolution. But inter– and intra–fractional variations in tissue and motion pattern call for up–to–date vector fields. Thus, the necessary MVFs have to fit to the acquired on–board CBCT data. With this in mind, one possible option is to estimate the MFVs based on the gated 4D CBCT images. However, a robust motion estimation is required because of the severe angular sampling artifacts of gated 4D CBCTs.

Possible solutions of MVFs obtained from an image–based motion estimation consist of two components due to the angular sampling artifacts. One part represents the patient motion and the other contains an artificial motion pattern induced by the co–registration of artifacts. Constraints like the cyclic breathing motion pattern have to be incorporated to remove the artificial motion pattern from the MVFs.

Here, we propose a robust motion estimation method that combines a cyclic registration approach with a model for

image artifacts due to angular sampling. For this purpose the artifacts are simulated within a second 4D image series free of patient's respiratory motion by using the 3D CBCT image. A prior image is generated to suppress the information about patient's motion included in the 3D CBCT image. Information from the second 4D image series is used to improve the robustness of motion estimation by an enhanced separation of patient's respiratory motion from the artificial one.

II. Materials and Methods

A. Simulations

Simulations were carried out for evaluation. The geometry of the On–Board Imager's® and TrueBeam's™ integrated kV imaging unit (Varian Medical Systems, Palo Alto, CA) is basis of the acquisition configuration. To simulate projection data we used a clinical CT reconstruction of a patient thorax as a phantom. We created realistic MVFs to simulate respiration.

The deformation intensity is direction–sensitive with a maximum of 2 mm in posterior–anterior and 20 mm in superior–inferior direction. The deformation is directly coupled to the RPM signal, which was set to be continuous with a rate of 15 respirations per minute. The field of measurement was extended to 46.5 cm by a shifted detector. The rotation speed was set to the maximum of 6° per second. Quantum noise was added to the simulated projections to obtain an image noise level of 60 HU in 3D CBCT images.

B. Phase–Correlated Feldkamp (PCF) Reconstruction

We use the well–known Feldkamp–Davis–Kress (FDK) filtered backprojection as our standard reconstruction algorithm [4] and shifted detector weighting is done according to reference [5]. Continuity and periodicity of respiratory motion allow to correlate each respiratory phase with the corresponding projection data. In this work the projections were associated by a retrospective phase gating. The cycle itself is subdivided into several subsets (bins) of finite length called phase windows.

To obtain gated 4D CBCT images a phase–correlated Feldkamp (PCF) reconstruction considers the relation between projection data p and respiratory phase by just using the projections associated to one phase (window) and discarding all the other projections. For an arbitrary phase bin n the corresponding operator is denoted as $\mathsf{X}^{-1}_{\mathrm{PCF}(n)}$ such that the respective PCF image $f_{\mathrm{PCF}(n)}$ is given by $f_{\mathrm{PCF}(n)} = \mathsf{X}^{-1}_{\mathrm{PCF}(n)}\, p$.

C. Motion–Compensated (MoCo) Image Reconstruction

In contrast to the subdivision into bins for gated 4D CBCT, all projection data are applied for motion–compensated (MoCo) image reconstruction. To prevent the high dose usage being at the expense of the temporal resolution, respiratory motion is compensated via motion vector fields (MVFs).

The relation between two arbitrary respiratory states r_i, r_j is described by $f_i = f_j \circ \mathrm{T}^i_j$, where $f_i = f_{\mathrm{PCF}(i)}$, $f_j = f_{\mathrm{PCF}(j)}$ are the respective phase volumes with $f = f(x, y, z)$, and T^i_j a transformation describing the MVF. Given a retrospective phase gating with N phase bins and also given that the transformations T^i_j are known for each phase pair $(i, j) \in N \times N$, the motion–compensated reconstruction $f_{\mathrm{MoCo}(i)}$ for an arbitrary respiratory phase bin $i \in N$ is described by

$$f_{\mathrm{MoCo}(i)} = \sum_{j=1}^{N} \left(\mathsf{X}^{-1}_{\mathrm{PCF}(j)}\, p \right) \circ \mathrm{T}^i_j .$$

I.e., to compensate for respiratory motion each single phase–correlated backprojection is warped by applying the motion vector fields T^i_j corresponding to the respiratory motion.

D. Robust Motion Estimation Using an Artifact Model

1) Cyclic Registration with Temporal Constraints: A cyclic registration method with temporal constraints [3] being based on the demons algorithm [6] is part of our robust motion estimation strategy. Due to a potential interference by image artifacts as it is the case for gated 4D CBCT images, a cyclic approach is used. There, motion is estimated between adjacent phases only and MVFs of non–adjacent phases are given by concatenation of MVFs from adjacent phases (c.f. figure 2).

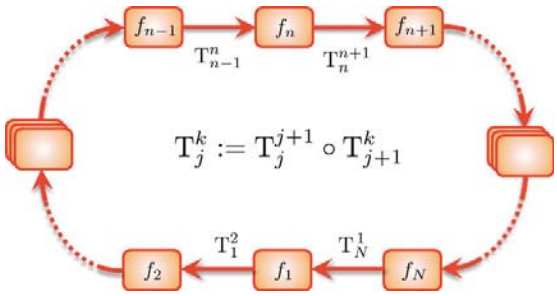


Figure 2. Illustration of the cyclic registration: f_j denotes the source images for the phase bins $j \in [1, N]$. The motion vector fields T^{j+1}_j describe the corresponding motion in–between two adjacent phase bins. The approach can also be applied in reverse order.

In addition to the cyclic approach, the assumption of a periodic breathing pattern is applied as temporal constraint by the cost function

$$E := \sum_{k=1}^{N} \|\mathrm{E}_k\|^2 := \sum_{k=1}^{N} \left\| \left(\prod_{j=k}^{N+k-1} \mathrm{T}^{j+1}_j \right) - \mathrm{Id} \right\|^2 ,$$

where $\prod$ denotes the noncommutative concatenation of several vector fields, i.e. $\prod_{j=1}^{N} \mathrm{T}^{j+1}_j = \mathrm{T}^2_1 \circ \mathrm{T}^3_2 \circ \ldots \circ \mathrm{T}^1_N$. Note that all indices are to be understood modulo N. Here, the deviation of each vector field concatenation and the expected identity mapping Id is penalized. The error or cost function E is minimized by the iterative scheme over k

$$\begin{aligned} j = k: \quad \mathrm{T}^{j+1}_j &\leftarrow \mathrm{T}^{j+1}_j - \frac{E_k}{N} \\ j \neq k: \quad \mathrm{T}^{j+1}_j &\leftarrow \mathrm{T}^{j+1}_j - \frac{E_k \circ \prod_{l=k}^{j-1} \mathrm{T}^{l+1}_l}{N}, \end{aligned}$$

that incorporates the additional information from the concatenation error vector fields E_k into the approximated motion vector fields T^{j+1}_j (more details in reference [3]).

2) Motion–Free Model for Angular Sampling Artifacts: While non–cyclic parts of the artificial motion induced by angular sampling artifacts are eliminated by a cyclic registration approach, cyclic parts may remain. The goal is to separate also the cyclic components of respiratory motion and artificial motion. For this purpose we create a second series of 4D images, called 4D artifact images, that contain similar angular sampling artifacts while being free of patient motion.

By using the 3D CBCT image to generate this series, it has to be considered that respiratory motion is included in the 3D CBCT image. To suppress the information about patient motion, we generate a prior image first, similar to that prior image in [7]. For this purpose, air regions, soft tissue regions, and bone regions have to be identified first. This is done by applying a simple thresholding with automatically chosen thresholds to segment the different types. Finally, the segmented air and soft tissue regions are set to the mean value of the particular region. Bone pixels keep their values, as they vary too much to properly model them with just one value.

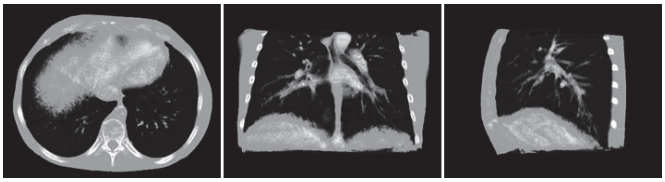


Figure 3. Prior image derived from 3D CBCT by simple thresholding to the segments air, soft tissue, and bone. Due to the presence of contrast agent some parts were classified as bone which were expected to be soft tissue.

The resulting prior image (c.f. figure 3) is forward projected in the same geometry as the rawdata were acquired. These simulated projection data are retrospectively binned by the same respiratory signal as the rawdata have been binned for the gated 4D CBCT. A phase–correlated Feldkamp reconstruction yields the 4D artifact images.

3) Robust Motion Estimation Using an Angular Sampling Artifact Model: Our proposed motion estimation method is illustrated as a flowchart in figure 4. Two separate registrations are performed using a cyclic registration approach, one for the gated 4D CBCT images on the one hand and one for the 4D artifact images on the other hand. The resulting vector fields from the gated 4D CBCT contain cyclic parts of both, patient's respiratory motion and angular sampling artifacts. Possible non–cyclic parts are eliminated by a cyclic registration as already mentioned. The results from the 4D artifact images contain only cyclic parts induced by the artifacts. The final step is to subtract these error approximations from the MVFs derived by gated 4D CBCT.

III. Results

PCF reconstructions and MoCo reconstructions were conducted for subsets with size $\Delta r = 10\%$ and with a step size of 5%, i.e. for 20 overlapping subsets.

Figure 5 shows mid–exhale phase bin views for different reconstruction techniques. Using the entire projection data without regard to respiratory phase 3D CBCT results in motion–induced blurring and streaking indicated by arrows.

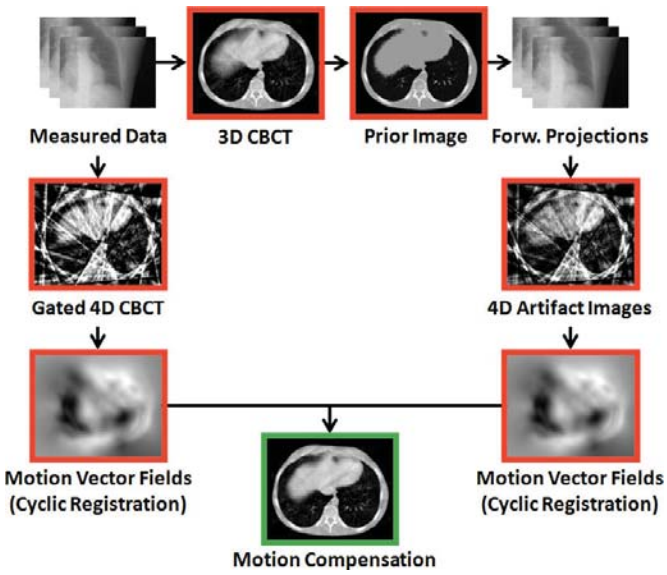


Figure 4. Illustration of the robust motion estimation method using an angular sampling artifact model: First a prior image is generated from the 3D CBCT by simple thresholding. A second 4D image series is derived by forward projection and gated reconstruction. For both 4D image series, gated 4D CBCT and 4D artifact images, MVFs are estimated using a cyclic registration approach. Finally, the results are combined and can be used for motion compensation.

Gated 4D CBCT comes with a high temporal resolution and reduced motion blurring. But its image quality is highly deteriorated by sparse–view artifacts like increased noise and prominent streak artifacts. Finer details like the pulmonary blood vessels cannot be identified.

To demonstrate, how important it is to consider the presence of sparse–view artifacts, we also show results for sMoCo, a standard motion compensation with conventional motion estimation. Here, the MVFs are obtained directly from their corresponding phase pair images applying a non–cyclic registration algorithm. Figure 5 shows sMoCo images that are strongly deteriorated by streaks and partially suffer from a low temporal resolution. In particular, lung details are blurred and impaired by streaks.

The results of aMoCo, a motion compensation using MVFs obtained from the artifact model–based motion estimation, show a similarly high temporal resolution like gated 4D CBCT. In contrast to gated 4D CBCT images the streaks are almost completely suppressed. Pulmonary blood vessels can be clearly seen in the aMoCo images.

To highlight also the importance of the artifact model, we performed a motion compensation (cMoCo) using MVFs obtained from cyclic registration directly. The results for both motion compensations, with and without using the angular sampling artifact model, are shown in figure 6. Cyclic parts induced by the angular sampling artifacts remain such that there are artifacts visible in cMoCo indicated by arrows, in particular in regions not affected by respiratory motion like the spinal region. These artifacts are suppressed by aMoCo due to the improved separation of cyclic patient motion from cyclic motion induced by angular sampling artifacts.

The so far qualitative results are supported by root mean square error (RMSE) measurements. Because of the heavy

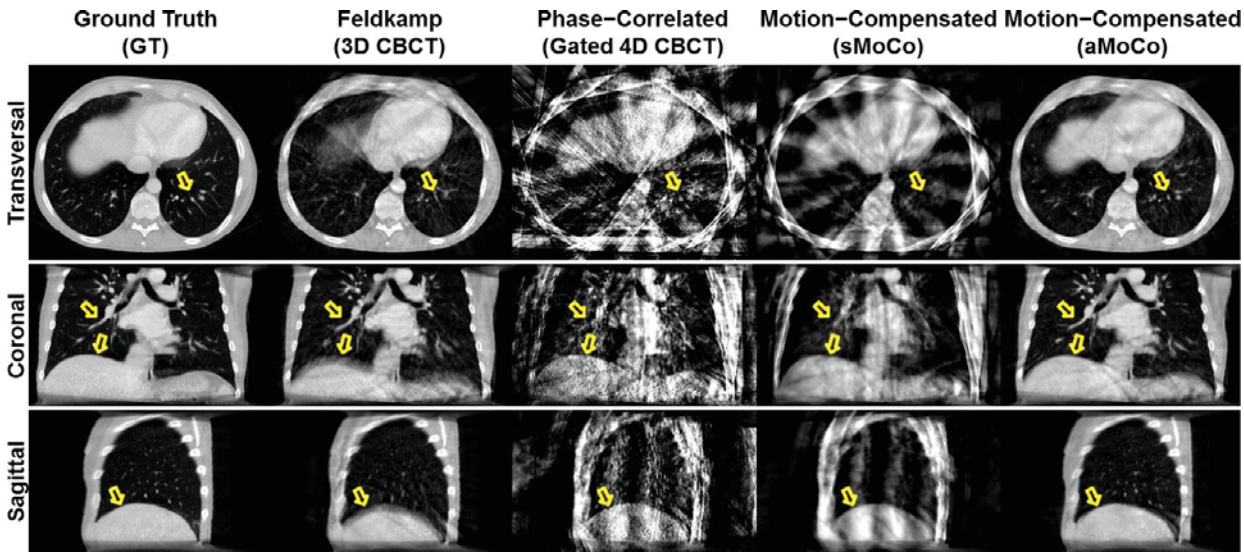


Figure 5. Ground truth and reconstructions of the simulated patient using different reconstruction techniques: A transversal (top row), a coronal (middle row) and a sagittal view (bottom row) of the mid–exhale phase bin is shown here. The ground truth (first column left), the standard Feldkamp reconstruction (second column left), the phase–correlated Feldkamp reconstruction (middle column) as well as the results of motion compensation using MVFs from a standard motion estimation (sMoCo, second column right) and from our proposed artifact model–based motion estimation (aMoCo, first column right) are shown for the simulation. All images are displayed at a grayscale window of $C = -200$ HU, $W = 1400$ HU.

streak artifacts the PCF reconstruction results in the greatest RMSE value of 473 HU. The visual impression of lowest motion blurring and lowest artifact impact in aMoCo images is confirmed by the lowest RMSE value of 122 HU compared to 183 HU for FDK and 156 HU for cMoCo.

IV. Summary and Discussion

We proposed a robust motion estimation method using an angular sampling artifact model to compensate for motion in case of on–board kV imaging units in radiation therapy. It separates patient's respiratory motion from cyclic and non–cyclic artificial motion parts induced by the angular sampling artifacts within the gated 4D CBCT images.

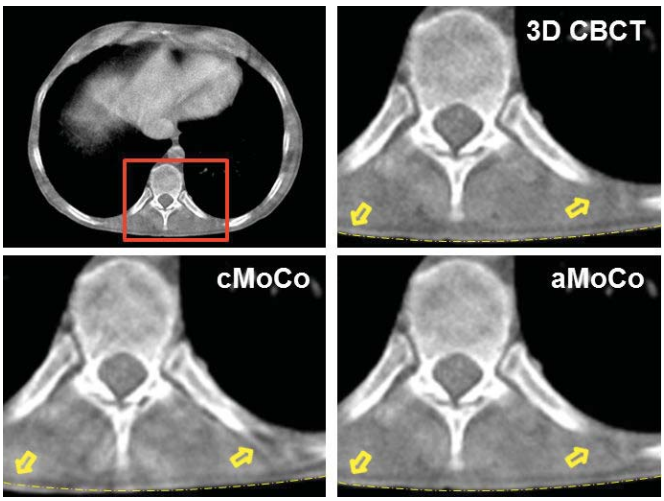


Figure 6. Increased robustness: A close–up of the spine region is shown for the 3D CBCT, and the results of motion compensation using MVFs from cyclic registration only (cMoCo) and from artifact model–based motion estimation (aMoCo). Artifacts within results of cMoCo are eliminated by aMoCo. All images are displayed at a grayscale window of $C = 0$ HU, $W = 1000$ HU.

This can be managed with the help of a cyclic registration with temporal constraints and an artifact motion model based on a series of second 4D image series, called 4D artifact images, which models the angular sampling artifacts of the gated 4D CBCT but free of patient's respiratory motion.

Based upon the model-based motion estimation a motion compensation aMoCo is possible which almost completely suppresses streak artifacts compared to the 4D CBCT images, while maintaining the high temporal resolution. The crucial part and therefore the enabling technique for motion estimation in the presence of image artifacts is the incorporation of the knowledge of a cyclic motion patterns and the incorporation of an artifact model. The temporal constraints are maintained by the model–based correction. Even more, the introduction of new artifacts is avoided by refining the motion vector fields instead of corrections on the volumes. Thus, the robustness of cyclic registration methods with temporal constraints is strengthened.

Acknowledgment

This work was supported by a research grant from Varian Medical Systems, Palo Alto, CA. Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] G. C. McKinnon and R. Bates, *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 2, pp. 123–127, Feb. 1981.
[2] F. Bergner *et al.*, *Med. Phys.*, vol. 36, no. 12, pp. 5695–5706, Dec. 2009.
[3] M. Brehm *et al.*, *Med. Phys.*, vol. 39, no. 12, pp. 7603–7618, Dec. 2012.
[4] L. Feldkamp *et al.*, *J. Opt. Soc. Am.*, vol. 1, no. 6, pp. 612–619, Jun. 1984.
[5] C. Maaß *et al.*, *Med. Phys.*, vol. 35, no. 12, pp. 5898–5909, Dec. 2008.
[6] J.-P. Thirion, *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, Sep. 1998.
[7] E. Meyer *et al.*, *Med. Phys.*, vol. 37, no. 10, pp. 5482–5493, Oct. 2010.

Four-dimensional projection-based motion estimation and compensation for cardiac x-ray computed tomography

Qiulin Tang, Jochen Cammin, and Katsuyuki Taguchi

***Abstract*—Our goal is to develop a fully four-dimensional computed tomography (CT) reconstruction algorithms for time resolved, low dose cardiac CT imaging. Toward this goal, two steps are required. First, cardiac motion is obtained by projection-based motion estimation, Second, motion compensated image reconstruction of a time-dependent deforming object from cone-beam projections is performed by utilizing the obtained cardiac motion.**

***Index Terms*—Computed tomography, motion estimation, motion compensation.**

I. Introduction

CARDIOVASCULAR disease is the leading cause of death in the western world, placing an ever-increasing burden on both private and public health services. The electrocardiogram (ECG)-gated cardiac computed tomography (CT) imaging is a promising non-invasive technique for early detection of fatty vulnerable plaque in coronary arteries. However, there are two major problems with the retrospectively ECG-gated image reconstruction technique: a large patient radiation dose and an insufficient temporal resolution, which results in motion artifacts in images.

Recently, algorithm-based approaches to reducing motion artifacts have been developed [1-6]. These algorithms have two main steps: motion estimation (ME) and motion compensated reconstruction (MCR) [2, 7]. Image-based ME was often utilized [5, 6], where an ECG-gated image reconstruction without motion compensation is performed for each cardiac phase, then image-based registration procedure is utilized to obtained the cardiac motion between cardiac phases. Then, the obtained cardiac motion is incorporated into the reconstruction process to compensate for the cardiac motion. One of the disadvantages of this type of algorithms is that the motion artifacts in ECG-gated reconstructions images could degrade the ME. An iterative algorithm that alternates ME and MCR is one of the solutions to this problem [6].

In the work, we proposed an alternative approach, projection-based ME, to avoid the degradation of estimated cardiac motion induced by motion artifacts. We hypothesize that the cardiac motion estimated by projection-based ME is more accurate than that estimated by image-based ME, in turn, reconstruction with less motion artifacts will be obtained by incorporating the more accurate cardiac motion.

The structure of this paper is as follows. In section II, we outline the motion estimation (ME)-motion compensated reconstruction (MCR) algorithm, the ME method, and MCR method. In section III, the proposed method is evaluated using a simulation based on a clinical cardiac CT data. Relevant issues are discussed in section IV, followed by conclusion.

II. Algorithms

A. Projection-based ME algorithm

The general framework of the proposed projection based ME method is similar to that of Zeng *et. Al.,* which was developed for respiratory motion estimation [8]. We have modified their approach to a weighted projection-based ME for cardiac imaging.

Deformation model:
In our method, cardiac motion is modeled by a linear combination of cubic B-spline basis functions. Let $f(\vec{x},\ t)$ be 4D image data, where $\vec{x}$ is the 3D spatial variable and t is the time variable. A 3D volume at time t_r is chosen as the reference volume, $f_r(\vec{x})$; 4D warping is then performed on the reference volume $f_r(\vec{x})$ to produce a warped 4D image $f_w(\vec{x},\ t)$:

$$f_w(\vec{x}', t) = f_r(\vec{x} + \vec{u}(\vec{x}, t), t_r), \quad (1)$$

where $\vec{u}(\vec{x}, t)$ is the 4D motion vector field (MVF). To decrease the number of unknowns, $\vec{u}(\vec{x}, t)$ is only sampled on knots $(\vec{\imath}|\Delta\vec{x}|,\ \tau\Delta t)$, where $|\Delta\vec{x}|$ and Δt are the knot spacing in the spatial domain and the temporal domain, respectively, and $\vec{\imath}$ and τ are discrete sampling indices in the spatial and temporal domains. Then, $\vec{V}(\vec{x}, t)$ at an arbitrary position is obtained by cubic B-spline interpolation. [8]

$$\vec{u}(\vec{x}, t) = \sum_{\tau\in T_0}\sum_{\vec{\imath}\in\vec{I}} \theta_{\vec{\imath},\tau}\, b\left(\frac{t}{\Delta t} - \tau\right)\beta\left(\frac{\vec{x}}{|\Delta\vec{x}|} - \vec{\imath}\right), \quad (2)$$

where $\theta_{\vec{\imath},\tau}$ is the cubic B-spline coefficient at knot postion, T_0 is 4 parameter indices in time, $\vec{I}$ is 4x4x4 parameter indices in space, b is a 1D cubic B-spline, and β is a 3D tensor product of cubic B-splines. Therefore, we obtain MVF $\vec{u}(\vec{x}, t)$ by obtaining cubic B-spline coefficients $\theta_{\vec{\imath},\tau}$. Therefore, we obtain MVF $\vec{V}(\vec{x}, t)$ by obtaining cubic B-spline coefficients $\theta_{\vec{\imath},\tau}$. A large number of knots may allow us to model MVFs accurately; however, it increases the computational cost and a risk of folding problem (or non-invertible MVF).

Cost function:
The noise in CT projections can be approximated by Gaussian noise; thus, a sum of squared weighted intensity

This work was supported in part by NIH grant R01 HL087918 and AHA grant BGIA 0865315E.

All of the authors are with The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore MD21287 USA (phone: 443-287-2974; fax: 410-614-1060; e-mail: ktaguchi@ jhmi.edu).

differences (SSWD) is used. The SSWD between measured projections $g(\vec{v},\beta)$ and projections of warped image $\mathcal{P}f_w(\vec{x},t_m)$ at time t_m is a function of the cubic B-spline coefficients $\theta_{\vec{\imath},\tau}$, and can be expressed as:

$$L_w\left(\theta_{\vec{\imath},\tau}\right) = \frac{1}{MN}\sum_{m=1}^{M}\left\|\mathcal{P}(w(\vec{x}))(g_m(\vec{u},\beta)-\mathcal{P}f_w(\vec{x},t_m))\right\|^2, \quad (3)$$

where N is the number of detector pixels, m is a time point, and M is the number of projections, $\mathcal{P}$ is projector, $w(\vec{x})$ is the weighting emphasize the region of interest, such of right coronary artery region.

We obtain cubic B-spline coefficients $\hat{\theta}_{\vec{\imath},\tau}$ by minimizing the following regularized SSWD cost function:

$$\hat{\theta}_{\vec{\imath},\tau} = \arg\min_{\theta_{\vec{\imath},\tau}} \Phi\left(\theta_{\vec{\imath},\tau}\right), \quad (4)$$

$$\Phi\left(\theta_{\vec{\imath},\tau}\right) = L_w\left(\theta_{\vec{\imath},\tau}\right) + R\left(\theta_{\vec{\imath},\tau}\right), \quad (5)$$

where $R(\theta_{\vec{\imath},\tau})$ denote a quadratic penalty term

$$R\left(\theta_{\vec{\imath},\tau}\right) = \frac{1}{2S}\alpha_x\sum_{\vec{\imath}}\sum_{\tau}\left|C_x\,\theta_{\vec{\imath},\tau}\right|^2 + \frac{1}{2S}\sum_{\vec{\imath}}\sum_{\tau}\alpha_t(\tau)\left|C_t\,\theta_{\vec{\imath},\tau}\right|^2, \quad (6)$$

where S is the number of 4D knots, α_x and $\alpha_t(\tau)$ are the weighting factors for the spatial and temporal smoothness terms, and C_x and C_t are differencing matrices in the spatial and temporal domains, respectively. It can be seen that weighting factors α_x and $\alpha_t(\tau)$ balance the contributions of similarity metric and the regularization term R in cost function Φ.

We can see that the regularization term R is cardiac phase-dependent and the similarity metric L is spatial-dependent. One can take advantage of it by applying a larger weight to an area with a complex motion pattern, which is difficult to be estimated.

Optimization:

The cost function $\Phi(\theta_{\vec{\imath},\tau})$ can be approximated by its Taylor expansion,

$$\Phi\left(\theta_{\vec{\imath},\tau}+\Delta\theta\right) = \Phi\left(\theta_{\vec{\imath},\tau}\right) + \Delta\theta^T\nabla\Phi\left(\theta_{\vec{\imath},\tau}\right) + \frac{1}{2}\Delta\theta^T\nabla^2\Phi\left(\theta_{\vec{\imath},\tau}\right)\Delta\theta + \cdots. \quad (7)$$

where $\Delta\theta^T$ is the transpose of vector $\Delta\theta$. The higher order terms are much smaller than the first three terms, and thus can be neglected:

$$\Phi\left(\theta_{\vec{\imath},\tau}+\Delta\theta\right) \approx c - \Delta\theta B + \frac{1}{2}\Delta\theta^T A\Delta\theta, \quad (8)$$

where

$$c = \Phi\left(\theta_{\vec{\imath},\tau}\right), \quad (9A)$$

$$B = -\nabla\Phi\left(\theta_{\vec{\imath},\tau}\right), \quad (9B)$$

$$A = \nabla^2\Phi\left(\theta_{\vec{\imath},\tau}\right). \quad (9C)$$

The matrix A, whose elements are the second partial derivatives of the function, is also called the Hessian matrix. B and A can be calculated by substituting Eq. (5) into Eq. (9B) and (9C), respectively,

$$B = -\frac{\partial\Phi}{\partial\theta_{\vec{\imath},\tau}} = \sum_{\tau=1}^{M}\frac{1}{MN}\left\langle \begin{matrix}\mathcal{P}w(\vec{x})\left(g_\tau(\vec{u},\beta)-\mathcal{P}f_w(\vec{x},t_\tau)\right),\\ \mathcal{P}w(\vec{x})\mathcal{P}_\tau\frac{\partial f_w}{\partial\theta_{\vec{\imath},\tau}}\end{matrix}\right\rangle - \frac{1}{S}\alpha_x C_x{}^T C_x\theta_{\vec{\imath},\tau} - \frac{1}{S}\alpha_t C_t{}^T C_t\theta_{\vec{\imath},\tau}, \quad (10A)$$

$$A = \frac{\partial^2\Phi}{\partial\theta_{\vec{\imath},\tau}\partial\theta_{\vec{\imath}'\tau'}} = \sum_{\tau=1}^{M}\frac{1}{MN}\left\langle \begin{matrix}\mathcal{P}w(\vec{x})\mathcal{P}_\tau\nabla f_r\, b\left(\frac{t}{\Delta t}-\tau\right)\beta\left(\frac{\vec{x}}{\Delta\vec{x}}-\vec{\imath}\right),\\ \mathcal{P}w(\vec{x})\mathcal{P}_{\tau'}\nabla f_r\, b\left(\frac{t}{\Delta t}-\tau'\right)\beta\left(\frac{\vec{x}}{\Delta\vec{x}}-\vec{\imath}'\right)\end{matrix}\right\rangle + \frac{1}{S}\alpha_x C_x{}^T C_x + \frac{1}{S}\alpha_t C_t{}^T C_t, \quad (10B)$$

where $\langle m,n\rangle$ denotes the inner product of arrays m and n.

The gradient of the cost function Φ at $\theta_{\vec{\imath},\tau}$ can be calculated by

$$\nabla\Phi\left(\theta_{\vec{\imath},\tau}\right) = -B + A\Delta\theta. \quad (11)$$

The cost function Φ is at an extremum if the gradient of Φ vanishes at $\theta_{\vec{\imath},\tau}$,

$$-B + A\Delta\theta = 0. \quad (12)$$

We use an iterative coordinate descent (ICD) approach to minimize the cost function; conjugate gradient (CG) method is chosen to minimize the cost of one MVF between the measured and calculated projection data. We chose ICD approach to handle a large number of unknowns efficiency, chose conjugate gradient because it often provides fast convergence and does not require an inversion of the Hessian matrix. From the current estimation θ^k, the next estimation θ^{k+1} can be obtained by the conjugate vector (searching direction) p^{k+1} as

$$\theta^{k+1} = \theta^k + \lambda^{k+1}p^{k+1}, \quad (13)$$

where λ^{k+1} and p^{k+1} can be calculated analytically using an approach similar to [6].

Implementation:

The proposed projection based motion estimation method can be implementation by following steps:

Step 1: Start motion estimation: $\theta_{\vec{\imath},\tau} = 0$. To save searching time, the MVFs estimated by image-based ME can be used as a start also.

Step 2: Generated estimated projections $\mathcal{P}f_w(\vec{x},t_m)$ as follows.

1: Choose a volume image of quiet phase $f_r(\vec{x})$, for example, mid-diastole, as a reference phase.

2: Perform 1D Bspline interpolation according to time on $\theta_{\vec{\imath},\tau}$ to obtain MVF ($16\times16\times16\times3000$ knots) for each time point t_m.

3: Perform 3D Bspline interpolation on sparse MVF to obtain dense MVF ($512\times512\times512$ knots) for each time point t_m.

4: Warped the reference image $f_r(\vec{x})$ with the dense MVF, and perform "dual source scanning" to generate projections, we call it measured projections.

Step 3: Calculate B and A using Eq. (9).

Step 4: Update the $\theta_{\vec{\imath},\tau}$ using Eq. (13).

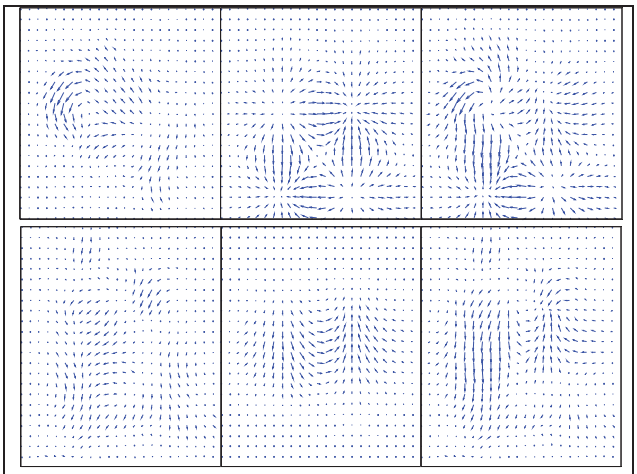
Step 5: Repeat step 2-4 till converged.

B. MCR algorithm

We used Schafer's motion tracking cone-beam backprojection method [2] and reconstructed volumes at cardiac phases t_m, $m = 1,\ldots,N_m$.

Schafer's method: It is the standard Feldkamp algorithm except that during the backprojection process, a ray that

corresponds to a new pixel location, $\vec{x}'(t) = \vec{x}(t_r) + \vec{V}(\vec{x}, t)$, will be chosen with the inverse squared-distance weight calculated by the new pixel location.



(a) (b) (c)

Fig. 1: MVFs: column (a) is the initial MVF, column (b) is the artificial MVF, and column (c) is the synthesized MVF. The upper row is the axial view and the lower row is the coronal view.

III. EVALUATIONS

We implemented the proposed algorithm using C and CUDA programming and a graphic processing unit board, and evaluated its performance with computer simulation.

A. Projection data

A patient image Data set:
A clinical cardiac volume image with less motion bluring and artifacts is used as a phantom, and it was obtained by a 64-slice CT scanner (Sensation 64; Siemens Healthcare, Forchheim, Germany) with a standard cardiac protocol, detector collimation of 2 × 32 × 0.6 mm with a z-flying focal spot technique, helical pitch of 0.29/rot., gantry rotation speed of 330 msec/rot, and 1160 projections/rot. ECG signals were also acquired during the scan. The volume size is 512 × 512 × 321, and the spacing was 0.43 mm × 0.43 mm × 0.50 mm.

Synthesized cardiac motion:
To simulation the cardiac motion, 4D motion vector fields (MVFs) were synthesized as follows [9]:
Step 1: The MVFs estimated by image-based motion ME for this patient were chosen as the initial MVFs $\vec{u}_0$ (see Fig. 1(a)). The initial MVFs modeled a realistic cardiac motion since they were estimated from the same patient data, although it may be degraded by motion artifacts.
Step 2: Additional MVFs, $\vec{u} = (u_x, u_y, u_z)$, were generated using Eq. (14) (also see figure 1(b)) considering reasonable cardiac motion such as expansion, contraction, and twisting.

$$u_x(x,y,z,t) = sin\left(\frac{2\pi(t-t_r)}{K}\right)\left(\frac{x}{10} + 2.2\,sin\left(\frac{3\pi x}{X}\right)\right)cos\left(\frac{3\pi y}{Y}\right) exp\left(-4\left(\frac{2x}{X}-1\right)^2 - 4\left(\frac{2y}{Y}-1\right)^2 - 4\left(\frac{2z}{Z}-1\right)^2\right), \quad (14A)$$

$$u_y(x,y,z,t) = sin\left(\frac{2\pi(t-t_r)}{K}\right)\left(\frac{y}{10} + 2.2\,sin\left(\frac{3\pi y}{Y}\right)\right)cos\left(\frac{3\pi x}{X}\right) exp\left(-4\left(\frac{2x}{X}-1\right)^2 - 4\left(\frac{2y}{Y}-1\right)^2 - 4\left(\frac{2z}{Z}-1\right)^2\right), \quad (14B)$$

$$u_z(x,y,z,t) = sin\left(\frac{2\pi(t-t_r)}{K}\right)\left(2.2\,cos\left(\frac{\pi z}{Z}\right)\right) exp\left(-4\left(\frac{2x}{X}-1\right)^2 - 4\left(\frac{2y}{Y}-1\right)^2 - 4\left(\frac{2z}{Z}-1\right)^2\right), \quad (14C)$$

Step 3: Synthesized MVFs $\vec{u}_s$ were obtained by summing $\vec{u}_0$ and $\vec{u}$. The synthesized MVFs were used as a true MVF (See Fig *1*(c)) in this study.

Forward projection:
Synthesized patient projection data for dual-source CT were generated using the patient image data set and the synthesized cardiac motion as follows. The patient image data at arbitrary time was obtained by warping the image data by the synthesized MVFs. Two projection data 90° apart were then generated by forward projecting. Therefore, two sets of projections were generated.
The parameters used for cardiac motion and scanner are: heart rate of 60 bpm, gantry rotation speed of 300 msec/rot, source to detector distance of 1000 mm, source to isocenter distance of 600 mm, 1000 projections/rot, and 3500 projections in total.

B. ME algorithm

The MVF was modeled using B-splines with 16 × 16 × 11 × 20 knots in the *x*, *y*, *z* in space and *t* in time, knot spacing were 14.6 mm, 14.6 mm, 16.0 mm, and 5% R-R interval (or 50 msec) respectively. To save computation time, during motion estimation, only one fifth projections were utilized during ME.
Figures 2 and 3 shows results of the motion estimation. A subtraction image (Figs. 2d and 3d) shows that the difference between the reference image f_0 and the image at the end-diastole (0%) was quite large. By warping the reference image with the estimated MVFs and repeating the subtraction (Figs. 2e and 3e), the differences were significantly decreased. This shows that the MVFs estimated by the proposed projection-based ME was qualitatively accurate. The means absolute difference between the estimated MVFs and the truth is 7.6%.

C. ME-MCR algorithm

There is still room to improve the accuracy of estimated MVFs (see Figs. 2e and 3e). We plan to perform MCR once the accuracy of ME is improved.

IV. DISCUSSION AND CONCLUSIONS

We have developed a fully four-dimensional image reconstruction method which consists of projection-based motion ME and MCR algorithms. We chose projection-based ME instead of image-based ME in this work. The merit of projection-based ME is that the data acquisition time for each projection view is very short (less than 1msec), and thus the cardiac motion during this short data acquisition can be ignored. In contrast, that of image-based ME is as long as 75 msec for dual-source CT, which results in motion artifacts that may degrade the accuracy of ME, especially for rapid

moving phase. The disadvantage of projection-based ME is that the motion along the projection ray is difficult to obtain accurately. The dual source CT, which acquires two projection data sets at view angles perpendicular to each other, does not have such a problem.

The proposed projection-based ME was evaluated by using synthesized patient data. The estimated MVFs appear to be qualitatively accurate and the means absolute error of the estimated MVFs was 7.6%.

We will continue the evaluation of the proposed method. By the time of the conference, we will perform quantitative assessment of the proposed projection-based ME and ME-MCR.

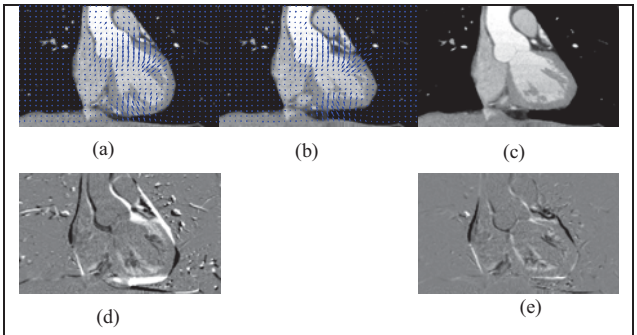


Fig. 2: The results of ME for patient. Coronal view is shown: (a) the estimated MVFs superimposed on the reference images at 75% of R-R waves of ECG signals ; (b) the estimated MVFs on the target images; (c) the warped reference images using the estimated MVFs; (d) the difference images, (a)-(b); and (e) the difference images, (c)-(b). The window width and level were 500 HU and 50 HU for columns (a)-(c) and 500 HU and 0 HU for (d)-(e), respectively.

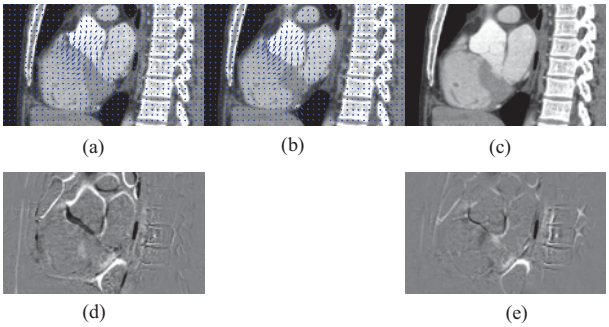


Fig. 3: The results of ME for patient. Sagittal view is shown. The 5-panel format and window settings are the same as in Fig. 2.

References

[1]C. Blondel, R. Vaillant, G. Malandain, and N. Ayache, "3D tomographic reconstruction of coronary arteries using a precomputed 4D motion field," Phys. Med. Biol. 49, 2197–2208 (2004).

[2]D. Schäfer, J. Borget, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational x-ray angiography," IEEE Trans. Med. Imaging 25, 898–906 (2006).

[3]A. A. Isola, A. Ziegler, T. Koehler, W. J. Niessen, and M. Grass, "Motion compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions," Phys. Med. Biol. 53, 6777–6797 (2008).

[4]C. O. Schirra, C. Bontus, U. van Stevendaal, O. Dössel, and M. Grass, "Improvement of cardiac CT reconstruction using local motion vector fields," Comput. Med. Imaging Graph. 33, 122–130 (2009).

[5]A. A. Isola, M. Grass, and W. J. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," Med. Phys. 37, 1093–1109 (2010).

[6]Q Tang, J Cammin, S Srivastava, and K Taguchi, "A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT," Med. Phys. 39(7), 4291–4305 (2012).

[7]K. Taguchi and H. Kudo, "Motion compensated fan-beam reconstruction for nonrigid transformation," IEEE Trans. Med. Imaging **27**(7), 907–917 (2008).

[8]R. Zeng, J. A. Fessler, and J. M. Balter, "Respiratory motion estimation from slowly rotating x-ray projections: Theory and simulation," Med. Phys. **32**(4), 984–991 (2005).

[9]Q Tang, S Srivastava, J Cammin, and K Taguchi, "Nonrigid four dimensional cardiac motion estimation" *SPIE* Medical Imaging 2012, San Diego, CA Feb. 4-9 (2012).

Opening Windows – Increasing Window Size in Motion-Compensated ECG-gated Cardiac Vasculature Reconstruction

Chris Schwemmer, Christopher Rohkohl, Günter Lauritsch, Kerstin Müller, and Joachim Hornegger

***Abstract*—In interventional angiographic C-arm CT imaging (rotational angiography), 3-D reconstruction of coronary vasculature is a topic of ongoing research. Due to the slow gantry rotation speed, motion artefacts corrupt image quality. Many approaches use retrospective ECG-gating to limit data inconsistencies during reconstruction. This poses a trade-off between gating window size and artefact level. A large gating window reduces undersampling artefacts, but increases motion artefacts and vice versa.**

In this paper, we investigate how motion compensation can be used to successively increase the gating window size in a bootstrapping process. We use a deformable 2-D–2-D registration between the acquired projection data and a forward projection of the previous reconstruction to estimate motion inside the current gating window. We evaluated the approach using the publicly available CAVAREV platform and on six human clinical datasets. We found that an increased gating window size leads to better homogeneity and resolution of fine detailed structures and a reduction of undersampling artefacts, while motion artefacts can be controlled well up to a gating window size of 80%, depending on speed and amplitude of the motion. In addition, the use of more projection data allows for a sharper ramp filter kernel, increasing the sharpness of the reconstructed structures. The CAVAREV results showed a 10% improvement over the best result published online at the time of this writing.

I. Introduction

During coronary interventions, three-dimensional information can provide improved guidance and easier assessment, especially for complex vessel topologies. For intra-procedural imaging, an angiographic C-arm CT system is a readily available modality. But the slow rotation speed of these devices limits their temporal resolution. This leads to motion-related artefacts like motion blur and streak artefacts. A retrospectively ECG-gated reconstruction of the X-ray projection data improves temporal resolution. Only images from a specific heart phase contribute to the reconstruction. However, this presents a trade-off regarding the gating window size. Projection images within a small gating window are expected to display a similar motion state. But the small amount of data in turn leads to undersampling artefacts that strongly decrease 3-D image quality. On the other hand, a large gating window avoids undersampling artefacts, but then residual motion within the gated projection data again leads to motion artefacts.

It has been shown in previous work that motion compensation can be used to correct for residual motion in ECG-gated reconstruction [1], [2]. These approaches have in common that first, a reference image is reconstructed, that is then used for the motion estimation. Since this image needs to show as little motion-related artefacts as possible to allow for a stable motion estimation, a smaller gating window is preferred. The resulting undersampling artefacts can be reduced by using a smooth ramp filter kernel, which unfortunately also reduces spatial resolution. But still, motion estimation for projection images far from the reference heart phase (large gating window) is difficult. Therefore, in this paper, we investigate whether and how motion estimation and compensation can be used to "bootstrap" a reconstruction with a large gating window and a sharper kernel in an iterative manner.

C. Schwemmer, K. Müller and J. Hornegger are with the Pattern Recognition Lab, Department of Computer Science, and the Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, email: chris.schwemmer@cs.fau.de. C. Rohkohl and G. Lauritsch are with the Siemens AG, Healthcare Sector, Forchheim, Germany.

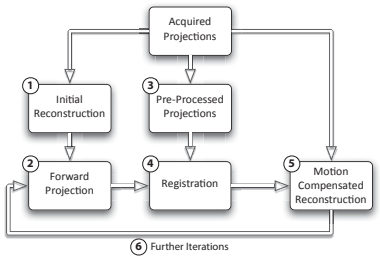


Figure 1: Illustration of our algorithm.

II. Methods

A. Motion Estimation and Compensation Algorithm

An overview of the motion estimation and compensation algorithm we used can be seen in Fig. 1. Most parts were published in [2], where a detailed description can be found. In the first step, an initial ECG-gated reconstruction is performed. In the second step, non-vascular tissue is removed by a thresholding operation. The vascular structure is forward projected using a maximum intensity forward projection. In step three, the original projection images are pre-processed using a morphological top-hat operation [1] and a thresholding, so that non-vascular tissue is also removed as much as possible. In the fourth step, the pre-processed original projections and the forward projections are registered using deformable 2-D–2-D registration in a multi-resolution scheme. In step five, a motion compensated, ECG-gated reconstruction is performed using the deformation field from the registration step. In the sixth step, the procedure is repeated for further refinement using the same or different gating parameters.

A set of parameters is available for our algorithm. All ECG-gated reconstructions are defined by the reference heart phase h_r and the size and shape of the gating window centred around

h_r. $h_r \in [0, 1]$ is expressed as a fraction of the heart cycle. The gating window is of $\cos^a$ shape [3], where $a \geq 0$ controls the edge steepness. The total size $\omega \in [0, 1]$ is given as a fraction of the heart cycle. For all gated reconstructions, streak reduction [4] was performed. Thresholding of the reconstructions before forward projection was performed by retaining only the t_r percentile of the largest voxel values. Thresholding of the original projections after top-hat filtering was performed by retaining only the t_p percentile of largest pixel values.

The current motion model is a combination of affine motion and deformable motion, where the latter is modelled by uniform cubic B-splines. We used only the affine part on the lower resolution levels and both parts on the higher levels. The B-spline model is parametrised by the number of control points c in each dimension. The cost function for registration was normalised cross-correlation and the optimisation was driven by a gradient descent method.

B. Bootstrapping Method

Since the initial reconstruction must be performed without any motion compensation, a small gating window (here: $\omega = 0.4$) is needed to avoid residual motion as much as possible. Still, remaining motion inside that window degrades image quality, which can be compensated by the algorithm described in Section II-A. A direct increase of ω in the first iteration is difficult for two reasons: Both residual motion and undersampling artefacts from the small window size limit the quality of the reference image, increasing the chance of misregistration during motion estimation. In addition, the amount of motion within the gating window increases with ω. A motion model with a larger c would then be desirable, which in turn decreases numerical stability. We therefore increase ω in an iterative fashion. For a certain ω, residual motion is compensated and the result used as a reference image for a new iteration with increased ω. Since a large number of parameter combinations is possible, we used CAVAREV to evaluate different choices and selected the best candidates for the final evaluation on clinical data.

C. Experimental Setup

1) CAVAREV*:* CAVAREV [5] is a publicly available platform for the evaluation of cardiac vasculature reconstruction algorithms. We used the cardiac motion-only dataset for our evaluation, assuming a strict breath-hold protocol. This dataset consists of 133 simulated projection images created from a software phantom that shows a thorax and contrasted left and right coronary arteries. Each projection image has a size of 960×960 pixels and an isotropic pixel size of 0.32 mm. Source-isocentre-distance was ~80 cm and source-detector-distance ~120 cm. The reconstructed 3-D volumes have an isotropic voxel size of 0.5 mm and a size of $98^3\,\mathrm{mm}^3$. The reference heart phase was selected as $h_r = 0.90$.

We created motion compensated reconstructions with 0 (initial reconstruction), 1, 2 and 3 iterations of our algorithm. Iterations 0 to 2 used $\omega = 0.4$, while in iteration 3 reconstructions with $\omega = 0.8$ and $\omega = 1.0$ were tested. In addition, each reconstruction was both done with a smooth and a normal kernel. After all experiments with 2 iterations, we selected the best scoring reconstruction for forward projection for the remaining experiments with 3 iterations.

Table I: Clinical datasets used for the evaluation. The number of projections correspond to 40%, 80% and 100% gating.

Dataset	3-D Img. Vol. [mm^3]	Heart Rate [bpm]	#Projs. Used
LCA1	$140^2 \times 101$	77 ± 0.1	53 / 106 / 133
LCA2	$152^2 \times 107$	58 ± 0.4	53 / 105 / 133
LCA3	$152^2 \times 114$	52 ± 0.7	54 / 106 / 133
RCA1	$152^2 \times 110$	68 ± 1.5	53 / 105 / 133
RCA2	$131^2 \times 109$	71 ± 2.1	53 / 105 / 133
RCA3	$143^2 \times 119$	54 ± 1.9	54 / 107 / 133

Table II: Motion model configuration for the experiments.

	Resolution Level		
	Low	Med	High
1. & 2. Iter.	affine	affine	+B-spline, $c = 6$
3. Iter.	affine	+B-spline, $c = 6$	+B-spline, $c = 12$

2) Human Clinical Datasets: Six human clinical datasets were used for the evaluation (cf. Table I): In LCA1, LCA2 and LCA3, a left coronary artery was imaged. The patient in dataset LCA2 had a total occlusion in the proximal part of the LAD, which means that no contrast agent reached the LAD beyond this point. In RCA1, RCA2 and RCA3, a right coronary artery was imaged. All datasets were acquired using a five second rotational angiography with selective contrast agent administration (1–2 ml/s) on an Artis zeego C-arm device (Siemens AG, Healthcare Sector, Forchheim, Germany). Source-isocentre-distance was ~80 cm and source-detector-distance ~120 cm. Each dataset consists of 133 projection images with a size of 1240×960 pixels and an isotropic pixel size of 0.308 mm. The reconstructed 3-D volumes have an isotropic voxel size of 0.5 mm. The reference heart phase was selected as $h_r = 0.75$ for all human datasets.

Again, we created motion compensated reconstructions with 0, 1, 2 and 3 iterations. A smooth kernel was used for iterations 0 to 2 and both kernels were tested for iteration 3, due to the results of the CAVAREV evaluation (cf. Section III-A). As in the CAVAREV experiments, a 40% gating window was used for iterations 0 to 2 and both 80% and 100% for iteration 3.

3) Common Parameters: Thresholding was performed at $t_r = 0.005$ and $t_p = 0.2$. The size of the morphological kernel for top-hat filtering was 3.85 mm. We employed a multi-resolution registration scheme with 3 levels. The motion model configuration for the different experiments is listed in Table II. The maximum number of optimisation steps on each level was set to 200 for the affine and 250 for the deformable registration. Optimisation was stopped if the gradient magnitude of the NCC was below $3 \cdot 10^{-4}$. For $\omega = 0.4$ and $\omega = 0.8$, a $\cos^4$ window was used, i.e. $a = 4$. For ungated reconstructions ($\omega = 1.0$), $a = 0$ was used. Streak reduction was used for $\omega = 0.4$ and $\omega = 0.8$.

D. Evaluation

Qualitative evaluation was carried out visually. The quantitative evaluation of the CAVAREV experiments was done using the metric $Q_{3D} \in [0, 1]$ provided by the platform [5], which describes the morphological similarity of a reconstruction to the ground truth data. $Q_{3D} = 1$ would indicate the best possible value.

Table III: CAVAREV results. The percentage is the size of the gating window ω.

	Q_{3D}
Initial, smooth kernel	0.744
Initial, normal kernel	0.739
1 Iter., 40%, smooth kernel	0.776
1 Iter., 40%, normal kernel	0.771
2 Iter., 40%, smooth kernel	0.776
2 Iter., 40%, normal kernel	0.773
3 Iter., 80%, smooth kernel	0.808
3 Iter., 80%, normal kernel	0.810
3 Iter, 100%, smooth kernel	0.805
3 Iter, 100%, normal kernel	0.821

For the quantitative evaluation of the human clinical datasets, we calculated the vessel sharpness [6] of continuous vessel segments along each reconstructed tree. We selected the same branch along the LAD and LCX of each left coronary dataset, and the main branch of each right coronary artery dataset. The average lengths of the selected branches were 198 mm (LAD for LCA1 and LCA3), 174 mm (LCX) and 183 mm (RCA). The LAD of dataset LCA2 could only be segmented for the first 79 mm due to the occlusion. Along each branch, sharpness measurements were taken with a spacing of 1 mm and the reported values are the average values of all measurements for that branch.

III. Results and Discussion

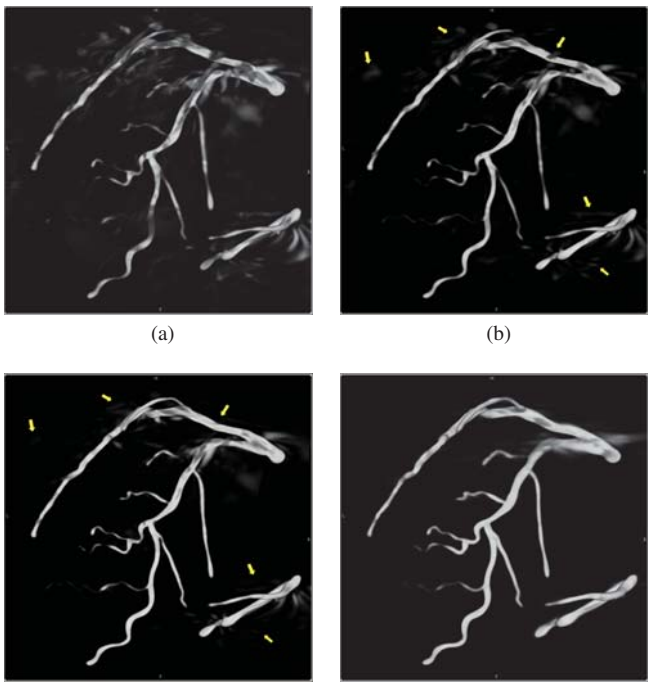
A. CAVAREV *Experiments and Parameter Selection*

Table III lists the Q_{3D} values for the CAVAREV experiments ($Q_{3D} = 0.744$ is the best value published online at the time of this writing). For this dataset, a second iteration with $\omega = 0.4$ and a smooth kernel does not change the result measurably. In addition, it can be seen that a smooth kernel leads to slightly better Q_{3D} values. From a theoretical viewpoint, $\omega = 0.4$ results in a low number of projections used for reconstruction, promoting undersampling artefacts. These are amplified by a sharper kernel. Therefore, we suggest a more conservative smooth kernel for both the initial and all motion compensated reconstructions that use a 40% gating window size. If a larger gating window is used, an improved reconstruction of the vasculature can be obtained, as shown by the higher Q_{3D} scores. Additionally, a sharper kernel does improve the achievable quality, since undersampling artefacts are not as dominant anymore.

Volume renderings of the reconstructions of selected parameter combinations can be seen in Fig. 2. Comparing Fig. 2b ($\omega = 0.4$) and 2c ($\omega = 0.8$), a clear decrease in artefact level can be observed as indicated by the arrows. In addition, vessel structures appear more homogeneous with a better visibility of distal parts. While an ungated reconstruction further improves vessel homogeneity, motion blur and an increase in artefact level can be observed in Fig. 2d.

B. Human Clinical Datasets

Table IV shows the vessel sharpness values for all datasets and reconstructions. Over all datasets, vessel sharpness decreased when going from a 40% to an 80% or 100%



(a) (b) (c) (d)

Figure 2: Reconstruction results of the CAVAREV dataset. (a) Initial reconstruction. (b) 2 iter., 40%, smooth kernel. (c) 3 iter., 80%, normal kernel. (d) 3 iter., 100%, normal kernel. The grey scale window was 1000 HU.

Table IV: Average vessel sharpness in 1/mm. The percentage is the size of the gating window ω, s.k. denotes smooth and n.k. normal kernel.

(a) Left coronary arteries.

	LCA1		LCA2		LCA3	
	LAD	LCX	LAD	LCX	LAD	LCX
Initial	0.410	0.368	0.324	0.354	0.453	0.423
1 Iter., 40%, s.k.	0.482	0.446	0.443	0.479	0.512	0.511
2 Iter., 40%, s.k.	0.486	0.464	0.451	0.484	0.524	0.516
3 Iter., 80%, s.k.	0.457	0.441	0.355	0.400	0.497	0.495
3 Iter., 80%, n.k.	0.550	0.523	0.498	0.543	0.633	0.589
3 Iter, 100%, s.k.	0.387	0.387	0.367	0.409	0.429	0.481
3 Iter, 100%, n.k.	0.451	0.456	0.467	0.528	0.537	0.555

(b) Right coronary arteries.

	RCA1	RCA2	RCA3
Initial	0.358	0.375	0.447
1 Iter., 40%, s.k.	0.500	0.457	0.483
2 Iter., 40%, s.k.	0.515	0.460	0.484
3 Iter., 80%, s.k.	0.481	0.436	0.451
3 Iter., 80%, n.k.	0.546	0.509	0.535
3 Iter, 100%, s.k.	0.451	0.410	0.427
3 Iter, 100%, n.k.	0.513	0.475	0.491

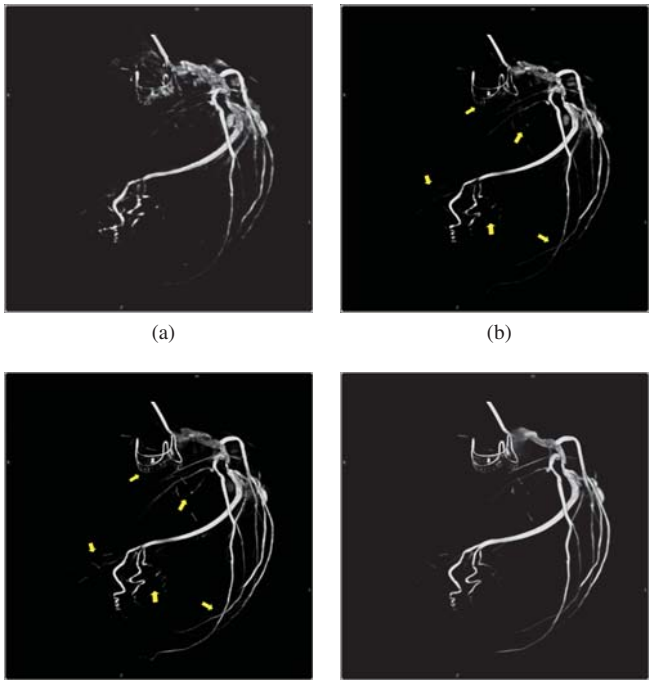


Figure 3: Reconstruction results of dataset LCA3 (left anterior oblique view). (a) Initial reconstruction. (b) 2 iter., 40%, smooth kernel. (c) 3 iter., 80%, normal kernel. (d) 3 iter., 100%, normal kernel. The grey scale window was 1000 HU.

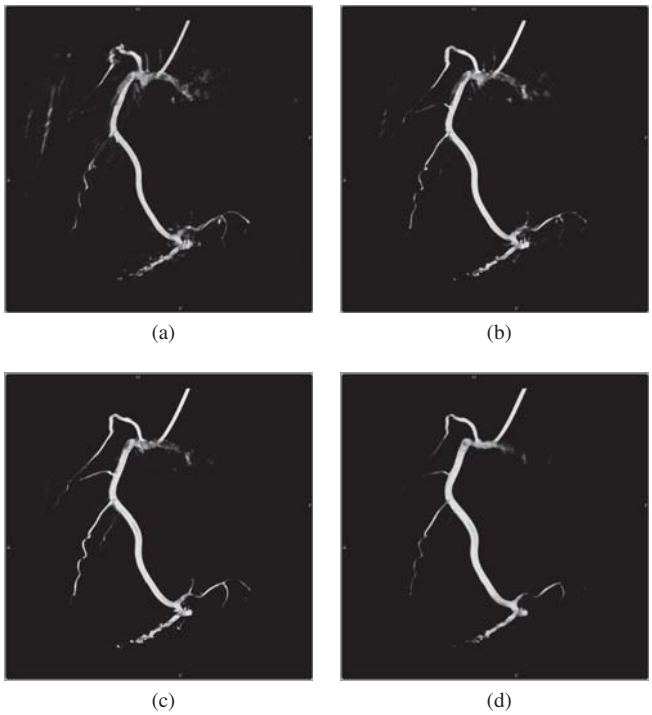


Figure 4: Reconstruction results of dataset RCA3 (left sagittal view). (a) Initial reconstruction. (b) 2 iter., 40%, smooth kernel. (c) 3 iter., 80%, normal kernel. (d) 3 iter., 100%, normal kernel. The grey scale window was 1000 HU.

gating window and a smooth kernel. But this effect can be compensated by switching to a sharper kernel: The best sharpness results for all datasets were achieved with $\omega = 0.8$ and a normal kernel.

In Fig. 3 and 4, reconstruction results for two datasets are shown as volume renderings. Fig. 3 illustrates the benefits of increasing the gating window size and being able to use a sharper kernel: Vessel homogeneity is greatly increased (cf. arrows), which in turn increases the visible length of small distal vessels. In addition, the depiction of the artificial valve is improved both in Fig. 3c and even more in Fig. 3d. While the same observations about vessel homogeneity and visibility hold for Fig. 4c, Fig. 4d shows that $\omega = 1.0$ did not improve but decrease vessel visibility for dataset RCA3 compared to $\omega = 0.8$. Again, we attribute this to not fully compensated motion.

IV. Conclusion

Gating window size in ECG-gated cardiac reconstruction is a trade-off between undersampling and motion-related artefacts. The latter can be reduced by residual motion compensation. But motion estimation and compensation becomes more difficult with large window sizes. We investigated how this can be overcome by an iterative process that successively increases the window size (bootstrapping). We found that motion-related artefacts can be controlled well up to a window size of 80%. The larger window reduces undersampling artefacts and leads to better homogeneity and resolution of fine detailed structures. In addition, more usable projection data allows for a sharper ramp-like filter kernel, which in turn increases sharpness and resolution of the reconstructions.

Acknowledgement

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative. The authors also would like to thank Prof. Dr. D. Böcker and Dr. P. Skurzewski, St. Marienhospital Hamm, Germany for acquiring clinical data.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] E. Hansis, D. Schäfer, O. Dössel, and M. Grass, "Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms," *Phys. Med. Biol.*, vol. 53, no. 14, pp. 3807–3820, 2008.

[2] C. Schwemmer, C. Rohkohl, G. Lauritsch, K. Müller, and J. Hornegger, "Residual Motion Compensation in ECG-Gated Cardiac Vasculature Reconstruction," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, F. Noo, Ed., Salt Lake City, UT, USA, June 2012, pp. 259–262.

[3] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography," *IEEE Trans. Med. Imaging*, vol. 25, no. 7, pp. 898–906, Jul. 2006.

[4] C. Rohkohl, G. Lauritsch, A. Nöttling, M. Prümmer, and J. Hornegger, "C-Arm CT: Reconstruction of Dynamic High Contrast Objects Applied to the Coronary Sinus," in *IEEE NSS/MIC Record*, Dresden, Germany, Oct. 2008, pp. M10–328.

[5] C. Rohkohl, G. Lauritsch, A. Keil, and J. Hornegger, "CAVAREV – An Open Platform for Evaluating 3D and 4D Cardiac Vasculature Reconstruction," *Phys. Med. Biol.*, vol. 55, no. 10, pp. 2905–2915, 2010.

[6] D. Li, J. Carr, S. Shea, J. Zheng, V. Deshpande, P. Wielopolski, and J. Finn, "Coronary Arteries: Magnetization-prepared Contrast-enhanced Threedimensional Volume-targeted Breath-hold MR Angiography," *Radiology*, vol. 219, no. 1, pp. 270–277, April 2001.

Towards General Motion Recovery in Cone-Beam Computed Tomography

Wolfgang Wein and Alexander Ladikos

***Abstract*—Irreproducible motion of either the patient or the device during a cone-beam X-Ray scan remains a major issue limiting reconstruction quality in many practical applications. Computational approaches are starting to emerge, which allow to model general motion parameters during the reconstruction itself. Besides, intelligent image processing on the projection data may reveal clues about "what went wrong" during a scan. We present a novel algorithm which uses a combined analysis in projection and reconstruction space, to both detect and account for unknown motion. This allows not only for the detection of large-scale, non-periodic bulk motion, but also an automatic recovery of it, required for a reconstruction void of artifacts. Using the proposed method, we can restore the reconstruction of clinical head scans with severe unknown motion. Moreover, we evaluate our method on synthetic data with known motion trajectories in a radiotherapy scenario.**

***Index Terms*—cone-beam, computed tomography, motion.**

I. Introduction

The cone-beam reconstruction algorithms used for most X-Ray Computed Tomography (CT) imaging devices today, strongly rely on the assumption that the geometry of the X-Ray source-detector arrangement relative to the imaged subject is correctly known and modeled at all times. However this assumption is often violated in light of non-reproducible device inaccuracies (which cannot be modeled during offline geometric calibration), as well as patient motion. The latter is more often an issue in slowly rotating C-arm systems as opposed to gantry-based CT scanners. Loss of resolution, severe artifacts, and completely wrong structures in the reconstruction appear, when the individual rays are considered with the wrong geometry. Ultimately it does not even matter if the excess motion stems from the device or the patient.

Prior work dealing with patient motion has generally focused on particular clinical applications or anatomic regions. This allows to constrain the problem in terms of motion characteristics, such as periodicity and typical trajectories occurring with cardiac or respiratory motion. If a surrogate signal for the motion phase is available, a binned 4D-reconstruction (with corresponding loss of signal to noise ratio proportional to the number of volumes), followed by further refinement [1] is possible. Other practical solutions require a reference scan not affected by motion, e.g. using a breath-hold acquisition protocol [2]. A more holistic approach is to jointly reconstruct the target volume and a motion model [3]. This however requires a reasonably close initialization of the motion parameters, hence unknown large-scale motion cannot be addressed. General mathematical formulations of such a joint reconstruction have been described e.g. in [4] and [5], with only early results on abstract geometric data though.

We had previously proposed a practical self-calibration approach for refining uncertain geometric and radiometric parameters within an algebraic reconstruction framework [6]. Other related methods also repeatedly reconstruct with altered parameters, using a quality criterion such as the sharpness of the reconstruction volume or a single slice thereof [7], in analogy to a camera's auto-focus. Such methods are able to tune parameters, however again require close initialization (if the initial reconstruction is deteriorated, a local optimization would not converge).

To address unknown large-scale motion during a scan, we have recently developed a method to analyze successive X-Ray images in projection space using the epipolar geometry [8]. It is able to detect and approximately recover the three-dimensional motion between X-Ray images with a small baseline (such as successive projections from a cone-beam scan). However simply concatenating those incremental motion estimates would yield drift of the entire sequence.

In this work, we extend the motion detection method [8] by appropriate normalization of the transformation estimates, such that they actually can be used to improve the reconstruction. We then combine the results with the self-calibration approach [6], in order to fine-tune the motion estimates. This yields a powerful hybrid technique which can fully recover even large-scale bulk motion. The remainder of this manuscript is organized as follows. First, we review the basics of self-calibration and motion detection in projection space. Then the new transformation normalization scheme allowing to incorporate the results into the reconstruction is presented. The combined method which yields the final refinement is then explained. Results on both real clinical and synthetic data sets are shown, followed by a brief discussion.

II. Methods

A. Self-Calibration

Algebraic reconstruction techniques generally minimize the re-projection error between a volume estimate and the measured X-Ray projection data:

$$E = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\| \tag{1}$$

where $\mathbf{A}$ is the system matrix, $\mathbf{x}$ the vector with all entries of the volume estimate, and $\mathbf{p}$ the X-Ray attenuation data. In a perfect scenario void of geometric errors, measurement noise, and with all details of the X-Ray physics modeled

ImFusion GmbH, Agnes-Pockels-Bogen 1, 80992 München, Germany. Corresponding author: Wolfgang Wein, E-mail: wein@imfusion.de.

in the forward- and back-projection steps, the residual re-projection error E would converge towards zero after sufficient iterations. It is therefore appropriate to minimize the same error E with respect to further unknown parameters (in addition to the individual attenuation values of the reconstruction volume). "Global" parameters which directly affect the entire reconstruction result, are optimized by re-computing a reconstruction, using equation 1 as cost function value to minimize. "Local" parameters such as additional drift or rotation of individual X-Ray projections, are sequentially optimized, after which the reconstruction is re-computed. The latter essentially comprises a 2D-3D registration algorithm of X-Ray projections to the reconstruction volume. A state of the art GPU implementation of an ordered subset simultaneous iterative reconstruction technique (OS-SIRT) is used, such that executing a non-linear optimization over the reconstruction as cost function is not computationally prohibitive [6].
For the "global" self-calibration, the number of parameters one can use is limited, both for increase in computation time and numerical stability (in terms of ambiguities and local optima). For the local sequential optimization of individual projection parameters, a close starting estimate is required, since otherwise the initial reconstruction would be deteriorated to start with. Therefore, these computational tools are in the current form rather suited for self-calibration of device parameters than for recovering large-scale patient motion.

B. Motion Detection

Using geometric calculations from stereo computer vision, it is possible to express the attenuation value $I(\mathbf{x})$ of an X-Ray image I at pixel location $\mathbf{x}$ as a linear combination of pixels along the epipolar line in the next X-Ray image J for a supposed geometric relationship between the images:

$$I(\mathbf{x}) = \sum_{k=1}^{n} \mathbf{w}_k J(\mathbf{x}_k) + \mathbf{w}_{n+1} \tag{2}$$

Here, $\mathbf{x}_k$ are a number of discrete sample locations along the epipolar line in image J, defined from preferred depth locations where the image structures are most prevalent. The unknown weights w_k are estimated locally within each pixel's neighborhood in a least-squares fashion. It is then possible to derive an image similarity measure for every image location from that:

$$S_x(I,J) = 1 - \frac{|\mathbf{i}_x - \mathbf{J}_x \mathbf{w}_x|^2}{Var(\mathbf{i}_x)} \tag{3}$$

where $\mathbf{i}_x$ is now a vector of pixel values within a neighborhood of I, $\mathbf{J}_x$ is a matrix with the number of pixels considered times the number of samples along the epipolar line, and $\mathbf{w}_x$ is the vector of weights computed around image location $\mathbf{x}$. Computed over entire pairs of successive X-Ray images, S constitutes a similarity measure which is sensitive to the supposed 3D-geometric relationship between the images. Varying the transformation between the images and hence modifying the epipolar line segments along which the image relationship is assessed, allows to detect if additional motion is present. In addition to the binary decision whether motion is present for every pair of X-Ray images, the relative motion between them can be approximately recovered by optimizing over the motion similarity [8].

C. Transformation Normalization

We describe the projection matrix of a cone-beam X-Ray frame i as

$$\mathbf{P}_i = \begin{bmatrix}\mathbf{K} & \mathbf{0}^\top\end{bmatrix} \mathbf{M}'_i \mathbf{R}_i \mathbf{M}_i \tag{4}$$

where the 4×4 matrix $\mathbf{R}_i$ describes the transformation from iso-center into detector coordinates (i.e. containing the rotation parameters of the cone-beam setup), and the 3×3 matrix $\mathbf{K}$ contains the intrinsic projection parameters. $\mathbf{M}'_i$ and $\mathbf{M}_i$ contain additional motion in detector and iso-center coordinates, respectively. Either of the latter ones can be optimized. For modeling patient motion, one would work in iso-center coordinates and describe the overall motion during a frame i as concatenation of incremental contributions from all previous frames:

$$\mathbf{M}_i = \prod_{k=1}^{i} \mathbf{T}_{i-k+1} \tag{5}$$

The relative transformations $\mathbf{T}_i$ computed by the motion detection algorithm for each single frame should add up to the identity matrix when closing the loop to the first frame (for a full 360° scan). In practice this is not the case due to drift and noise in the motion estimation. We therefore need a method for loop closing which takes the structure of the relative transformations into account but modifies them in such a way that they result in the identity transformation when concatenated.

More formally let us assume that we have n relative transformation matrices $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_n$ and a final relative transformation $\mathbf{C}$. We then want to find $\widehat{\mathbf{C}}_i$ so that

$$\begin{aligned}\mathbf{T} = \mathbf{C}\prod_{i=1}^{n}\mathbf{T}_{n-i+1} &= \left(\prod_{i=1}^{n}\widehat{\mathbf{C}}\right)\left(\prod_{i=1}^{n}\mathbf{T}_{n-i+1}\right) \\ &= \prod_{i=1}^{n}\widehat{\mathbf{C}}_{n-i+1}\mathbf{T}_{n-i+1}\end{aligned} \tag{6}$$

$\widehat{\mathbf{C}}$ is the n-th matrix root of $\mathbf{C}$. Assuming $\mathbf{C}$ is of the form

$$\mathbf{C} = \begin{bmatrix}\mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1\end{bmatrix}; \quad \widehat{\mathbf{C}} = \begin{bmatrix}\widehat{\mathbf{A}} & \widehat{\mathbf{t}} \\ \mathbf{0} & 1\end{bmatrix} \tag{7}$$

and $\mathbf{A}$ is diagonalizable, we can compute the eigenvalue decomposition of $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ and write $\widehat{\mathbf{C}}$ as above with

$$\widehat{\mathbf{M}} = \mathbf{V}\mathbf{D}^{\frac{1}{n}}\mathbf{V}^\top \tag{8}$$

$$\widehat{\mathbf{t}} = \left(\left(\sum_{i=1}^{n-1}\left(\prod_{j=1}^{i}\widehat{\mathbf{A}}\right)\right) + \mathbf{I}\right)^{-1}\mathbf{t} \tag{9}$$

Since $\mathbf{D}$ is a diagonal matrix $\mathbf{D}^{\frac{1}{n}}$ is obtained by simply taking the n-th root of each entry on the diagonal.

Finally, we need to move the $\widehat{\mathbf{C}}$ matrices into the coordinate frame of each transformation $\mathbf{T}_i$. The $\widehat{\mathbf{C}}_i$ in $\mathbf{T} = \prod_{i=1}^{n}\widehat{\mathbf{C}}_i\mathbf{T}_i$ resulting from this coordinate system change are given by

$$\widehat{\mathbf{C}}_i = \mathbf{T}_i\mathbf{G}^{-1}\widehat{\mathbf{C}}\mathbf{G}; \quad \mathbf{G} = \prod_{j=1}^{n-i+1}\mathbf{T}_{n-j+1} \tag{10}$$

For the transformation chain given in equation 5 the incremental transform $\mathbf{C}$ we correct is $\mathbf{C} = \mathbf{M}_i^{-1}$.

D. Hybrid Refinement

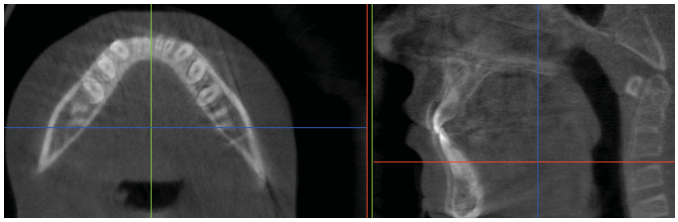
Our new overall algorithm operates as follows.
1) All projection images containing motion wrt. the previous frame are detected (including motion from the first frame to the last). This relative motion is successively found by optimizing over rigid transformation parameters for the detected frames. In the case of a full scan, the scheme described in section II-C is used to normalize the optimized transformations in order to avoid drift. A reconstruction after this step generally exhibits significantly reduced motion artifacts.
2) The main blocks of connected frames without motion are selected. Rigid parameters for the transformation of all blocks but the largest (which remains fixed) are fed into the global self-calibration algorithm. Before every evaluation of parameters, the transformations of all frames outside of blocks are normalized according to the interpolation scheme described in section II-C, in order to maintain the initial estimate from the motion compensation. This step usually yields the largest overall improvement, since the rough motion estimation on projection images is optimized using the residual error of the reconstruction, eliminating drift but keeping the overall structure of inter-frame motion.
3) A local self-calibration step (i.e. 2D-3D registration) is executed for all frames outside of blocks, hence refining transformation parameters in $\mathbf{M}_i'$ wrt. detector coordinates. An improved reconstruction is computed after the optimization of all frames. This step may be repeated once or twice, until the residual error does not improve over a selected ϵ threshold.
4) Optionally, a local self-calibration step on all frames is conducted. This would also ensure a final compensation of slight device inaccuracies.

Since both the motion similarity measure S and the self-calibration residual error E yield smooth cost function values, an optimization algorithm which internally approximates derivatives results in a significantly lower number of evaluations than other direct search methods. We therefore use the Bound Optimization by Quadratic Approximation (BOBYQA) method [9].

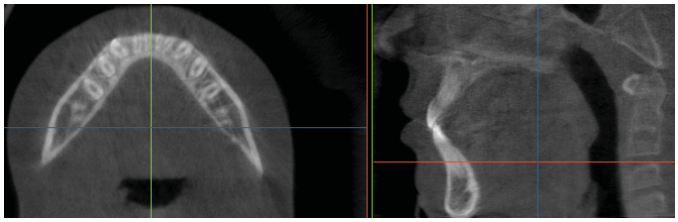
III. Results

A. Clinical Data

We have applied our algorithm on several orthodontic head scans with strong patient motion, however without ground truth information available. Figure 1 depicts axial and sagittal cross-sections of a reconstruction before and after our motion recovery method. In this example, the patient has moved his head several times during the sequence, with steady phases in between (which is addressed by the block optimization in step 2 of our method). The reduction of artifacts is clearly visible, in particular in the wrong "shadow" of the front teeth in the original reconstruction. This data set is a full scan with 450 frames and excentric detector motion for enlarged field of view. Five blocks of consistent successive frames have been found in this sequence, hence the block-based self-calibration in step 2 takes $(5-1)*6 = 24$ parameters. The non-linear optimization terminates after about 250 cost function evaluations, a single evaluation takes about 0.5 seconds with reconstruction volume dimension 256, on a NVIDIA GeForce GTX 670 GPU. All steps combined, therefore including the local self-calibration of all frames, yield a computation time in the order of five minutes. The following table shows the residual errors averaged over all frames after each step.



(a) original reconstruction



(b) after motion recovery

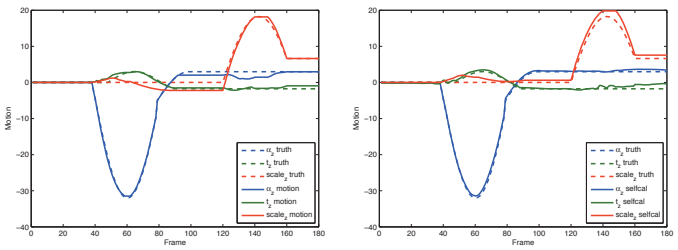
Fig. 1. Reconstruction result on a real patient before and after motion recovery

step	error	step	error
original	0.00983316	step 3	0.00900554
step 1	0.00926058	step 3 repeated	0.00899732
step 2	0.00903790	step 4	0.00888388

Those values are expressed in average absolute differences per pixel, with intensities normalized to $[0 \dots 1]$.

B. Synthetic Data

We also evaluated our method on a thorax CT scan with simulated patient motion, by generating 180 DRRs on a 360 degree trajectory around the patient. The artifical motion included rigid bulk motion as well as affine non-uniform scaling to resemble respiratory motion, with the resting states after motion being different than before (see dotted lines in figure 2). Running the projection-based motion optimization we were able to approximately recover the motion parameters (see figure 2a). Running the self-calibration step on these parameters further reduced the error particularly in frames between motion, as seen in figure 2b. Figure 3 shows an axial and a sagittal slice through the reconstructed patient volume (resolution 512^3) at different stages of the motion optimization together with the residual error of the reconstruction. The final reconstruction quality is close to the ground truth, both in terms of visual appearance and residual error. The average per-frame spatial error for the center point of the reconstruction volume and its direction were reduced from 6.53 ± 9.40mm to 0.66 ± 0.59mm in translation, and from $1.29 \pm 0.90^\circ$ to $0.55 \pm 0.34^\circ$ in rotation.



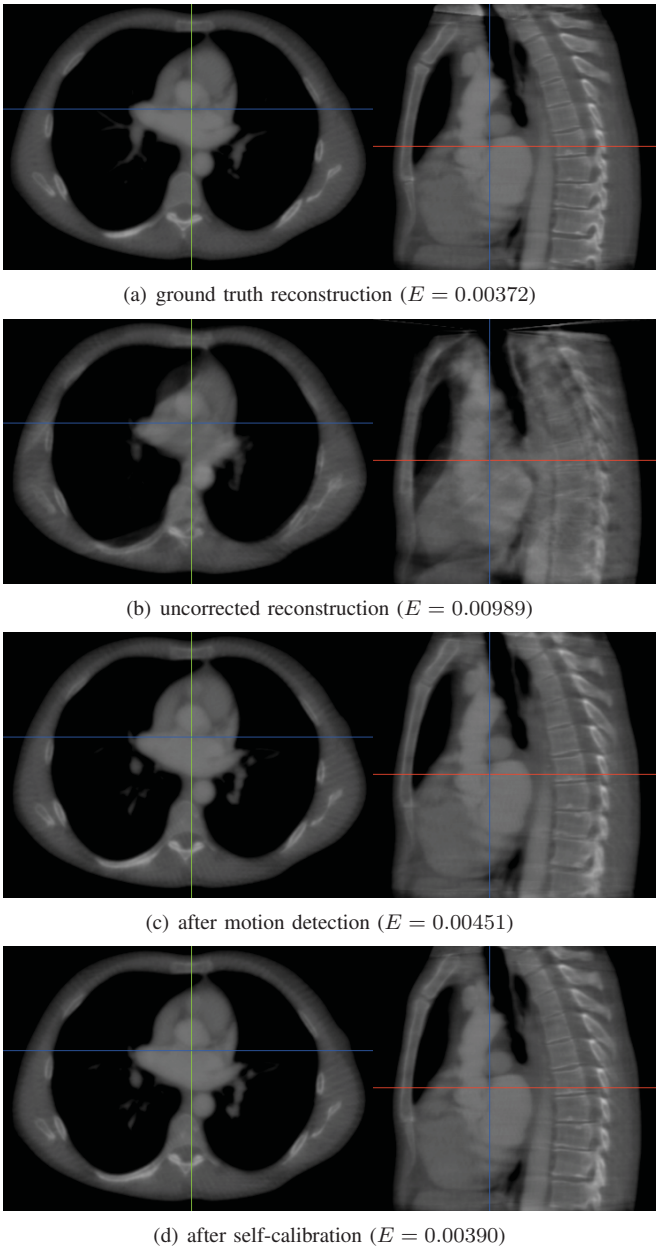
(a) after motion detection (step 1) (b) after-self calibration (step 4)

Fig. 2. Comparison of ground-truth and recovered patient motion

IV. Discussion

We have developed a new hybrid approach operating in both projection and reconstruction space to detect and recover patient or device motion in cone-beam CT. Detection of large motion is achieved using the projection based motion analysis, and approximately estimated with it. Only in combination with a self-calibration approach can the motion be precisely recovered however. The link between the intermediate results of relative motion between frames, and absolute motion parameters that are optimized using global block based self-calibration, is achieved by a normalization of the chain of transformation matrices.

We have shown that this method can restore severely compromised head scans, bootstrapping the motion information without any prior assumptions. To the best of our knowledge, this has not been achieved before. While these qualitative clinical results were obtained using a rigid parametrization, linear affine transformation matrices are supported as well (as demonstrated in section III-B), and should allow to deal with a majority of clinical scenarios involving respiratory motion. In order to tackle complex motion with significant local deformations (i.e. cardiac motion), adapted parametric motion models can be integrated with our method. For a given parameter configuration, it can then be evaluated simultaneously in projection and reconstruction space, how well it describes the actual motion. While this first approach suggested here is rather straightforward, we believe that such a joint analysis is a powerful foundation for dealing with motion in general. More complex motion models are the subject of future work, as well as thorough experiments on clinical data with ground truth (e.g. by tracking actual patient motion with auxiliary sensors). Further work is also required for a more theoretical understanding of the connection between the epipolar geometry in successive X-Ray images with intermediate results during reconstruction.



(a) ground truth reconstruction ($E = 0.00372$)

(b) uncorrected reconstruction ($E = 0.00989$)

(c) after motion detection ($E = 0.00451$)

(d) after self-calibration ($E = 0.00390$)

Fig. 3. Reconstruction results and residual errors without and with motion recovery on simulated data.

References

[1] C. Rohkohl, G. Lauritsch, L. Biller, and J. Hornegger, "ECG-gated interventional cardiac reconstruction for non-periodic motion," in *MICCAI 2010 Proceedings, Part I, LNCS 6361*, 2010, pp. 151–158.

[2] R. Zeng, J. Fessler, and J. Balter, "Estimating 3-D respiratory motion from orbiting views by tomographic image registration," *IEEE Trans. Med. Imag.*, vol. 26, pp. 153–163, Feb. 2007.

[3] E. Hansis, H. Schomberg, K. Erhard, O. Dössel, and M. Grass, "Four-dimensional cardiac reconstruction from rotational x-ray sequences - first results for 4D coronary angiography," in *SPIE Medical Imaging 2009 Conference*, vol. 7258, 2009.

[4] S. Brandt and V. Kolehmainen, "Motion without correspondence from tomographic projections by bayesion inversion theory," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2004, pp. 582–587.

[5] M. Lyksborg, M. Hansen, and R. Larsen, "A statistical approach to motion compensated cone-beam reconstruction," in *IEEE International Symposium on Biomedical Imaging*, 2010, pp. 804–807.

[6] W. Wein, A. Ladikos, and A. Baumgartner, "Self-calibration of geometric and radiometric parameters for cone-beam computed tomography," in *Fully3D 2011 Proceedings*, Jul. 2011.

[7] Y. Kyriakou, R. Lapp, L. Hillebrand, D. Ertel, and W. Kalender, "Simultaneous misalignment correction for approximate circular cone-beam computed tomography," *Physics in Medicine and Biology*, vol. 53, pp. 6267–6289, 2008.

[8] W. Wein and A. Ladikos, "Detecting patient motion in projection space for cone-beam computed tomography," in *MICCAI 2011 Proceedings*, ser. Lecture Notes in Computer Science. Springer, Sep. 2011.

[9] M. J. D. Powell, "The BOBYQA algorithm for bound constrained optimization without derivatives," Department of Applied Mathematics and Theoretical Physics, Cambridge England, Tech. Rep. NA06, 2009.

TOF ML-ACF reconstruction using CT based attenuation as *a priori* information

Vladimir Y. Panin, Michel Defrise and M.E. Casey

***Abstract*– In modern PET-CT scanners attenuation information is acquired by CT machinery. This information may not be fully compatible for use in PET data correction. The source of PET and CT inconsistency is due to the sequential nature of scanning, a shorter CT FOV, and scans of differing durations.**

Recent theoretical investigations concluded that both activity and attenuation distributions can be obtained from PET emission TOF data alone up to knowledge of the sinogram scaling parameter. In TOF, attenuation is recovered in the form of attenuation factors. This allows for development of a fast iterative algorithm, referred to as ML-ACF. In this work we consider the combination of attenuation information obtained from a CT scan and from TOF PET data for PET data correction. We constrained ML-ACF attenuation factor estimations by biasing them toward CT information by means of a quadratic prior. Preliminary results showed that the activity image can gain noise improvement while being free of artifact, induced by mismatched CT attenuation.

I. Introduction

TOF PET scanners were commercially introduced in the past decade, exploiting advances in scanner hardware. The use of additional TOF information led to decreased noise and therefore potentially better detectability of small lesions [1] in studies of large patients (relative to the TOF resolution). It was observed that reconstructed images were less affected by the bias in the estimated correction factors, such as normalization [2], scatter estimation [2], and attenuation [3, 4]. In broader terms, the use of TOF information led to better definition of the consistent part of the data, regardless of the nature of inconsistencies, which may be due to high frequency statistical data noise or to lower frequency artifacts in scatter estimation.

Most data correction factors are specific to the physics of PET and to the characteristics of the scanner. Therefore, these factors should ideally be estimated based on the PET data themselves and on the scanner type. The current practice, however, requires using an external device to estimate the attenuation correction. In modern PET-CT scanners, attenuation information is provided by the attenuation map (converted to 511 keV energy) obtained from a CT scan. In the past, numerous attempts were made to recover attenuation information from PET data alone. In the non-TOF case, only limited successes were achieved because it was difficult to avoid cross-talk between the activity and the attenuation images. Cross-talk artifacts contaminate the reconstructed activity image with features present in the attenuation map images and vice versa.

Recent theoretical investigations [5] concluded that both activity and attenuation distributions can be determined from PET TOF data up to a global scaling of the activity image. One limitation of this approach is that the attenuation correction factors cannot be determined outside of the support of the emission sinogram. This limitation complicates the estimation of the scatter background, which requires knowledge of the full attenuation map. In addition, the stability of the simultaneous estimation of emission and attenuation will need more investigation, in particular to determine its dependence on the TOF resolution.

Fortunately, in the current practice of PET imaging, the estimation of attenuation from PET data alone is not necessary due to the diagnostic value of PET-CT scanning. Nevertheless, the available CT-based attenuation information is not fully compatible with the PET data. The CT field-of-view is shorter and yields only truncated patient data. In addition, the PET data are acquired under different conditions. The most noticeable mismatch between the attenuation map and activity image is due to respiratory motion. In Fig. 1(a) CT image was taken while the patient was holding his breath. Therefore, the patient diaphragm was in the lowest position. During PET acquisition, the liver region was averaged over respiratory movement. This mismatch between PET and CT is clearly seen in the activity image in Fig. 1(b), which was reconstructed using the usual OS-EM algorithm. This image shows the liver split in two parts, only the lower of which was properly corrected for attenuation.

The theory [5] was formulated in terms of an attenuation factors estimation based on analytical consistency conditions. This result suggests that a similar approach, based on estimating the attenuation factors rather than the attenuation image, should also be possible for a discrete model based on the maximum likelihood (ML) estimation. This led to the TOF ML-ACF algorithm [6-8], which has the same complexity as the commonly used ML-EM algorithm. The reconstruction in Fig. 1(c) was obtained by applying two iterations of ML-ACF, starting with the result of one iteration of the OS-EM reconstruction. After these two ML-ACF iterations, the liver and spleen activity images were significantly changed.

The first attempts for simultaneous ML estimation of activity and attenuation [9-11] were performed by estimating the attenuation map rather than the attenuation factors. In this case, typical values of the attenuation coefficients can be used as prior information, solving the problem of the arbitrary scale factor. These approaches also provide additional constraints on the attenuation factors, which can lead to better and less noisy estimates of the activity. Scatter estimation can be performed based on the current estimate of the attenuation map. In

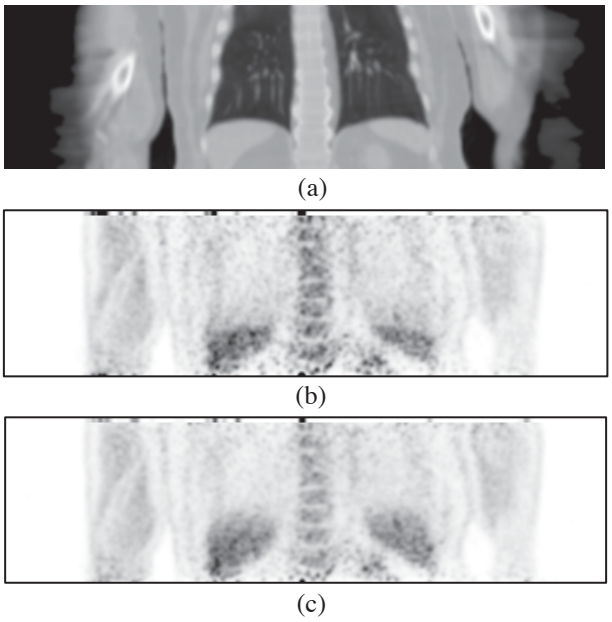


Figure 1. Reconstruction of patient data. (a) Attenuation map, which is a smoothed version of the CT image; (b) activity image after the OS-EM 3 iterations, 24 subsets where CT based attenuation map was used for attenuation correction; (c) activity image after ML-ACF 2 iterations, 24 subsets, where 1st iteration of OSEM was used as initial condition. The activity images were post-smoothed by 4 mm Gauss filter.

contrast, the ML-ACF algorithm has the advantage of fast reconstruction, but it lacks the above mentioned possibilities. In this paper, we propose to overcome this limitation by combining ML-ACF with the use of available, but potentially incomplete or inaccurate, CT information on a PET-CT scanner. The combination is based on a simple quadratic prior, which penalizes large differences between the estimated attenuation factors and those derived from the CT scan.

II. Methods

A. Objective Function

TOF prompt data $y=\{y_{it}\}$ with spatial projection (LOR) index i and TOF bin index t, can be modeled by combining the modeled projection $\bar{p}$ from the emission object f, corrected for scanner efficiency by a known normalization array N and for attenuation by an array of attenuation factors a. The background events have a known mean $\bar{b}$, equal to the sum of the estimated, efficiency corrected, scatter, and of the estimated randoms. The Poisson Likelihood objective function has the following form:

$$L(\mathbf{f},\mathbf{a})=\sum_{i,t}\left(y_{it}\ln\left(a_i\bar{p}_{it}+\bar{b}_{it}\right)-a_i\bar{p}_{it}-\bar{b}_{it}\right),$$
$$\bar{p}_{it}=\sum_j\frac{c_{it,j}}{N_i}f_j \qquad (1).$$

where $c_{it,j}$ is the system matrix corresponding to the TOF geometric projector. In the following, quantities without index t denote quantities summed over the TOF index, e.g.

$$c_{i,j}=\sum_t c_{it,j}\ ,\ y_i=\sum_t y_{it}\ ,\ \bar{p}_i=\sum_t\bar{p}_{it}\ ,\ \bar{b}_i=\sum_t\bar{b}_{it} \qquad (2).$$

The available CT based attenuation is accommodated in two ways. The attenuation of LORs, passing outside the support of the activity image, cannot be recovered from emission data. In the rest of the paper, prime spatial projection indices denote such LORs, for which attenuation values are provided by CT and are not estimated. These LORs can be defined by CT attenuation sinogram values, which are close to one. For the other LORs, the attenuation values are estimated from the emission data, but are biased towards the CT values by a quadratic prior. We now describe two algorithms.

B. Nested loop algorithm

This algorithm alternatively updates a and f. We use the following objective function:

$$L(\mathbf{f},\mathbf{a})-\beta U(\mathbf{a})=\sum_{i',t}\left(y_{i't}\ln\left(a_{i'}^{CT}\bar{p}_{i't}+\bar{b}_{i't}\right)-a_{i'}^{CT}\bar{p}_{i't}-\bar{b}_{i't}\right)+$$
$$\sum_{i,t}\left(y_{it}\ln\left(a_i\bar{p}_{it}+\bar{b}_{it}\right)-a_i\bar{p}_{it}-\bar{b}_{it}\right)-\beta\sum_i\left(a_i-a_i^{CT}\right)^2 \qquad (3)$$

where $\beta>0$ is a regularization parameter.

This objective function is maximized with respect to a, assuming fixed $\bar{p}$, using a surrogate based iteration, where the logarithmic term in (3) is replaced by the same separable surrogate as in the derivation of the ML-EM algorithm. The resulting update step is:

$$a_i^{(n+1)}=\frac{2C_i^{(n)}}{B_i+\sqrt{B_i^2+4AC_i^{(n)}}}$$
$$A=2\beta,\ B_i=\bar{p}_i-2\beta a_i^{CT},\ C_i^{(n)}=\sum_t\frac{y_{it}\bar{p}_{it}a_i^{(n)}}{a_i^{(n)}\bar{p}_{it}+b_{it}} \qquad (4).$$

Instead of this iteration, it is possible to directly obtain a closed form solution if one uses data precorrected for background. In that case, b is zero in (4), the factor $a^{(n)}$ is simplified, hence

$$C_i=\tilde{y}_i\equiv smooth(y_i-b_i) \qquad (5),$$

and (4) directly yields the maximizer a of (3) in terms of the current activity estimate f.

After updating the attenuation, the objective function (3) is maximized with respect to f at fixed a, using the ML-EM algorithm. Typically-ordered subsets are used and a single iteration of OS-EM is done instead of a full maximization.

C. Attenuation by-product algorithm

In the absence of regularization ($\beta=0$), the update (4) using the pre-corrected data (5) becomes $a_i=\tilde{y}_i/\bar{p}_i$. We will refer to this solution as the Gaussian approximation, since it maximizes the Gaussian approximation of the Poisson Likelihood, with data variance approximated by $\bar{p}$. We proposed this method in our previous work [8], as a way to incorporate background events in ML-ACF without requiring

an alternative nested-loop update strategy because substituting $a_i = \tilde{y}_i / \bar{p}_i$ into (3) leads to a likelihood that can easily be maximized. In order to accommodate a similar approach we minimize a Gaussian approximation of (1) with a modified quadratic prior,

$$\sum_{it} \frac{1}{\bar{p}_{it}} \left(a_{it} \bar{p}_{it} - \tilde{y}_{it} \right)^2 + \beta \sum_i \bar{p}_i \left(a_i - a_i^{CT} \right)^2 \qquad (6).$$

The prior is chosen to allow a closed form expression for the attenuation factors at fixed activity:

$$a_i = \frac{1}{1+\beta} \left(\frac{\tilde{y}_i}{\bar{p}_i} + \beta a_i^{CT} \right) \qquad (7).$$

This expression has valid limits for zero β and large β.

Substituting (7) into the un-regularized Poisson likelihood (1), the following objective function must be maximized:

$$L = L' + \sum_{i,t} y_{it} \ln\left(\tilde{y}_i \bar{p}_{it} + (1+\beta) \bar{b}_{it} \bar{p}_i + \beta a_i^{CT} \bar{p}_{it} \bar{p}_i \right) - \sum_{it} y_{it} \ln\left(\bar{p}_i \right) - \frac{\beta}{1+\beta} \sum_i a_i^{CT} \bar{p}_i \qquad (8).$$

where L' stands for the contribution of the primed indices (first line in (3)). A separable surrogate for L is derived as in [6, 8, 12], with the difference that Jensen's convexity inequality is applied to the first log term in (8) by decomposing its argument as a sum of J+J^2 terms, with the J^2 terms for $\bar{p}_{it} \bar{p}_i \sim \sum_{jj'} c_{it,j} c_{i,j} f_j f_{j'}$. Maximizing the surrogate yields the update equation (including primed indices for LORs with fixed CT attenuation), which is:

$$f_j^{(n+1)} = \frac{f_j^{(n)}}{\sum_{i'} \frac{c_{i',j} a_{i'}^{CT}}{N_{i'}} + \sum_i \frac{c_{i,j}}{N_i} \left(\frac{y_i}{\bar{p}_i^{(n)}} + \frac{\beta}{1+\beta} a_i^{CT} \right)} \times \left\{ \sum_{i',t} \frac{c_{i't,j}}{N_{i'}} \frac{y_{i't}}{\bar{p}_{i't}^{(n)} + a_{i'}^{CT-1} \bar{b}_{i't}} + \sum_{i,t} \frac{c_{it,j}}{N_i} \frac{y_{it} \left(\tilde{y}_i + \beta a_i^{CT} \bar{p}_i^{(n)} \right)}{\tilde{y}_i \bar{p}_{it}^{(n)} + (1+\beta) \bar{b}_{it} \bar{p}_i^{(n)} + \beta a_i^{CT} \bar{p}_{it}^{(n)} \bar{p}_i^{(n)}} + \sum_{i,t} \frac{c_{i,j}}{N_i} \frac{y_{it} \left((1+\beta) \bar{b}_{it} + \beta a_i^{CT} \bar{p}_{it}^{(n)} \right)}{\tilde{y}_i \bar{p}_{it}^{(n)} + (1+\beta) \bar{b}_{it} \bar{p}_i^{(n)} + \beta a_i^{CT} \bar{p}_{it}^{(n)} \bar{p}_i^{(n)}} \right\} \qquad (9).$$

In the limit of zero β, (9) tends to the ML-ACF used in [8]. However, the limit of algorithm (9) for large β does not coincide with ML-EM.

III. Results

The investigations were conducted on the patient data from Fig. 1 (where Fig. 1(c) was the ML-ACF reconstruction with the attenuation by-product algorithm of section II.C). The patient was scanned for 5 mins per bed position after injection with 293 MBq FDG. Only one bed position out of a total of eight was reconstructed here. The patient diaphragm was in the middle of the scanner. We used the standard mCT sinogram size of 400(radial) x168(azimuthal) x621(axial) x13(TOF). The reconstructed images are 400x400x109 2x2x2 mm size. The TOF resolution of the mCT is 570 ps with 312ps TOF bin width.

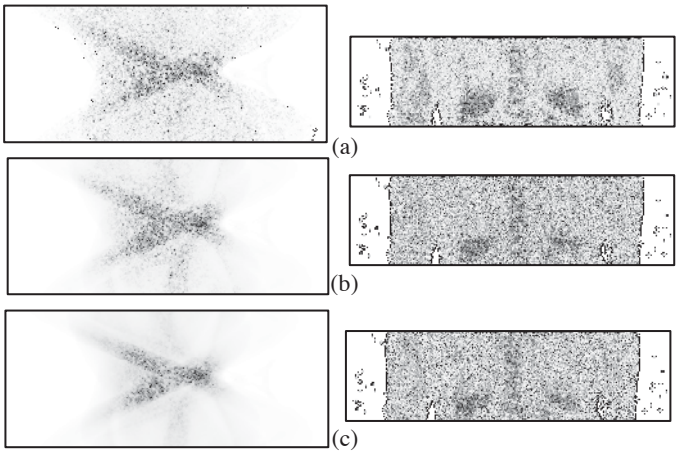


Fig. 2. Reconstruction by nested loop algorithm. The images on the left are estimated ACF sinograms (a^{-1}). The images on the right are activity estimations after 2 iterations, 24 subsets, 4 mm Gauss filter post-smoothing. (a) β=0; (b) β=100 and (c) β=500. The initial estimate was the OS-EM reconstruction with CT based attenuation, 1 iteration, 24 subsets.

We were trying to find the regularization parameter range that is the turning point in the definition of liver structure between OS-EM and non-regularized ML-ACF reconstructions. Fig. 2 represents images obtained with the nested loop algorithm of section II.B. Increasing the regularization parameter resulted in a worse definition of the liver, while the estimated attenuation sinogram became noticeably smoother. We measured noise in one planar ROI of the liver region. This plane was not affected by the change in attenuation due to ML-ACF, since it was low enough to be within the liver during the CT scan. Noise decreased by 7.2% in Fig. 2(b) and by 13.8% in Fig. 2(c), compared with the activity image of Fig. 2(a).

Fig. 3 presents similar observations for the attenuation by-product algorithm. Here the noise effect was less understood. The attenuation by-product algorithm is less noisy compared to the same number of iterations of the OS-EM algorithm, see Fig. 1. This was inherited by the regularized by-product algorithm. Even in the case of significantly large β attenuation, the by-product algorithm was smoother when compared to the OSEM reconstruction. Nevertheless, Fig. 4 displays that the regularized attenuation by-product algorithm became similar to the OS-EM reconstruction on a voxel-to-voxel basis. The voxel values inside the patient (an attenuation map defined patient boundaries; axial planes, which contain OSEM liver artifacts, were excluded) were considered in a histogram of absolute difference in each voxel, normalized by the OSEM reconstruction voxel value. The histogram became progressively sharper around zero difference with a β increase. The attenuation by-product images were scaled to match the global scaling of the OSEM reconstruction. A significantly

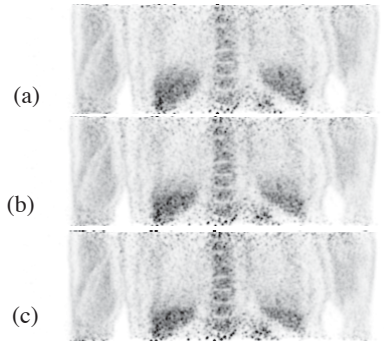


Fig. 3. Reconstruction by attenuation by-product algorithm. The activity images are after 2 iterations, 24 subsets, 4 mm Gauss filter post-smoothing. (a) β=0; (b) β=1 and (c) β=1000. The initial estimate was the OS-EM reconstruction with CT based attenuation, 1 iteration, 24 subsets.

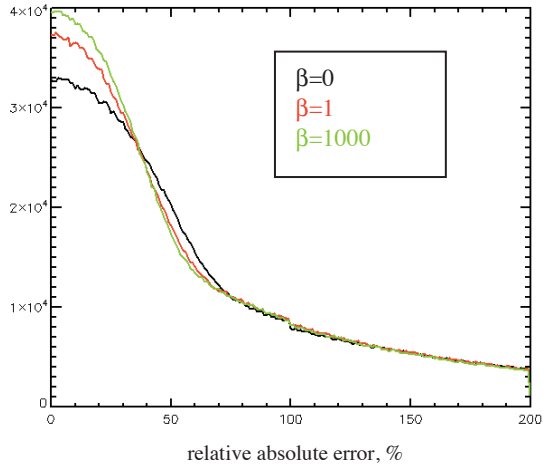


Fig. 4. Histogram of normalized absolute differences between voxel value of Fig. 3 reconstructions and voxel value of OS-EM reconstruction with CT based attenuation, 3 iterations, 24 subsets.

regularized image, β=1000, was identical in scaling to the OSEM reconstruction, while non-regularized reconstruction was different by 10%. The resolution-noise tradeoff of the attenuation by-product algorithm needs to be investigated to understand if the attenuation byproduct algorithm is effectively noisier when compared to the gold standard, the OSEM reconstruction.

IV. Discussion and Conclusions

Two algorithms for incorporating CT information into TOF simultaneous activity and attenuation factors reconstruction were presented. The preliminary results suggest that the a priori CT information decreases noise in the reconstructed activity image, compared to the use of only emission data. At the same time, allowing the attenuation factors to deviate from the CT-based factors limits the effect of the mismatch between sequential data acquisitions.

References

[1] C. Lois, B. W. Jakoby, M. J. Long, K. F. Hubner, D. W. Barker, M. E. Case, M. Conti, V. Y. Panin, D. J. Kadrmas, and D. W. Townsend, "An Assessment of the Impact of Incorporating Time-of-Flight Information into Clinical PET/CT Imaging," *J. Nucl. Med*, vol. 51, pp. 237-245, 2010.

[2] W. Wang, Z. Hu, E. E. Gualtieri, M. J. Parma, E. S. Walsh, D. Sebok, Y.-L. Hsieh, C.-H. Tung, X. Song, J. J. Griesmer, J. A. Kolthammer, L. M. Popescu, M. Werner, J. S. Karp and D. Gagnon, "Systematic and Distributed Time-of-Flight List Mode PET Reconstruction," *2006 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, San Diego, CA, M04-2, pp. 1715-1722, 2006.

[3] T.G. Turkington and J.M. Wilson, "Attenuation artifacts and time-of-flight PET," *IEEE Nucl. Sci. Symp. and Med. Imag. Conf. (Orlando, FL)* pp. 2297-2299, 2009.

[4] M. Conti, "Why is TOF PET reconstruction a more robust method in the presence of inconsistent data?" *Phys Med Biol*. vol. 56. pp. 155-168, 2011.

[5] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Phys. Med. Bio*., vol. 56, pp. 885-899, 2012.

[6] M. Defrise, A. Rezaei and J. Nuyts, "Analysis of the convergence of the MLACF algorithm", *The 12th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2013.

[7] J.L. Nuyts, A. Rezaei, and M. Defrise, "ML-Reconstruction for TOF-PET with Simultaneous Estimation of the Attenuation Factors," *2012 IEEE Nucl. Sci. Symp. Med. Imag. Conf*., M04-1, 2012.

[8] V.Y Panin, M. Defrise, J.L. Nuyts, A. Rezaei, M.E. Casey, "Reconstruction of uniform sensitivity emission image with partially known axial attenuation information in PET-CT scanner," *2012 IEEE Nucl. Sci. Symp. Med. Imag. Conf*., M04-7, 2012.

[9] A. Rezaei, J. Nuyts, M. Defrise, G. Bal, C. Michel, M. Conti and C.C. Watson, "Simultaneous Reconstruction of Activity and Attenuation for Time-of-Flight PET," *IEEE Trans. Med. Imag*., vol. 31, pp. 2224-2233, 2012.

[10] A. Salomon, A. Goedicke, B. Schweizer, T. Aach, and V. Schulz, "Simultaneous Reconstruction of Activity and Attenuation for PET/MR," *IEEE Trans. Med. Imag*., vol. 30, pp. 804-813, 2011.

[11] V.Y. Panin, M. Aykac, and M.E. Casey, "Simultaneous reconstruction of emission activity and attenuation coefficient distribution from TOF data, acquired with rotating external line source," *2011 IEEE Nucl. Sci. Symp. Med. Imag. Conf*., M5-202, pp. 4329 - 4336, 2011.

[12] A.R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag*., vol. 12, pp. 328–33, 1993.

Alternating Direction Method of Multiplier for Emission Tomography with Non-Local Regularizers

Se Young Chun, Yuni K. Dewaraja, and Jeffrey A. Fessler

***Abstract*—The ordered subset expectation maximization (OSEM) algorithm provides a fast image reconstruction method for emission and transmission tomography such as SPECT, PET, and CT by approximating the gradient of a likelihood function using a subset of projections instead of using all projections. However, for computationally expensive regularizers such as patch-based non-local (NL) regularizers, OSEM does not help much to improve the speed of reconstruction because one evaluates the regularizer gradient for every subset. We propose to use variable splitting to separate the likelihood term and the regularizer term for penalized emission tomographic image reconstruction problem and to optimize it using the alternating direction method of multiplier (ADMM). This new scheme allows us to run more sub-iterations for the optimization related to the likelihood term. We evaluated our ADMM for 3D SPECT image reconstruction with the patch-based NL regularizer that uses the Fair potential. Our proposed ADMM improved the speed of convergence substantially compared to other existing methods such as gradient descent, EM and OSEM using De Pierro's approach, and the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) algorithm.**

***Index Terms*—ordered-subset expectation-maximization, non-local regularizer, emission tomography, alternating direction method of multiplier**

I. Introduction

Statistical image reconstruction methods such as the expectation-maximization (EM) algorithm can improve quality of images for emission tomography such as PET and SPECT as compared to the analytical image reconstruction such as the filtered back-projection (FBP) [1]. It started to be used widely in clinics and in commercial PET and SPECT scanners after the fast algorithm called ordered-subset expectation-maximization (OSEM) was developed [2]. By approximating the gradient of a likelihood function using the subset of projections instead of using all projections, OSEM algorithm performed faster image reconstruction. This approximation has been used for unregularized emission tomographic image reconstruction [2] and regularized emission and transmission tomographic image reconstruction using simple quadratic or edge-preserving regularizers. Since the computation cost for these regularizers is fairly low compared to that for the likelihood term, the OSEM algorithm could often also speed up penalized likelihood (PL) image reconstruction.

Recently, patch-based non-local (NL) regularizers have been proposed that improve image quality compared to other conventional regularizers such as quadratic or edge-preserving functions in general image processing [3], PET reconstruction [4], and MRI reconstruction [5]. The same principle has been used for emission image reconstruction or super resolution using high resolution CT or MRI side information [6]–[9]. For emission tomography problems such as [4], [7]–[9], many optimization algorithms were used for image reconstruction such as the gradient descent (GD) [8], the EM (or OSEM) algorithm from the optimization transfer using De Pierro's lemma [4], the EM algorithm using one-step late approach [9], and the quasi-Newton algorithm called the Limited-memory Broyden-Fletcher-Goldfarb-Shanno with a box constraint (L-BFGS-B) [7]. Since the computation cost of the NL regularizers is very high compared to that of the likelihood, the OS does not help much to improve the convergence rate of PL image reconstruction.

In this paper, we propose to use variable splitting to separate the likelihood term and the regularizer term for penalized emission tomographic image reconstruction problem and to optimize it using the alternating direction method of multipliers (ADMM). This new scheme allows us to run more sub-iterations for the optimization related to the likelihood term. There have been some methods to use a variable splitting for the data fidelity term and the regularizer term [3], [10]–[12]. However, these previous methods split the variable to deal with non-smooth regularizers such as the total variation and to solve the sub-problem related to the regularizers using efficient methods such as shrinkage. Our proposed variable splitting has different motivation. We divide the original optimization into a few sub problems and we update the sub problem related to the NL regularizer less often.

We evaluated our new ADMM for 3D SPECT image reconstruction with a patch-based NL regularizer that uses the Fair potential [4]. Our XCAT phantom-based simulation [13] shows that our proposed ADMM improved the speed of convergence substantially compared to existing methods such as GD, EM and OSEM using De Pierro's approach, and the L-BFGS-B algorithm.

II. Method

A. Statistical image reconstruction for emission tomography

Statistical image reconstruction methods for emission tomography yield better image quality than non-iterative algorithms. The usual form of statistical image reconstruction is

This work was supported in part by NIH grant 2RO1 EB001994.

Se Young Chun was with the University of Michigan, Department of EECS and Radiology, Ann Arbor, MI 48109-2122, USA, and is now with Ulsan National Institute of Science and Technology (UNIST), Department of ECE, Ulsan, South Korea. (e-mail: delight@umich.edu, sychun@unist.ac.kr).

Yuni K. Dewaraja is with the University of Michigan, Department of Radiology, Ann Arbor, MI 48109-2122, USA. (e-mail: yuni@umich.edu).

Jeffrey A. Fessler is with the University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. (e-mail: fessler@umich.edu).

to perform the following constrained optimization with respect to an image $\boldsymbol{f}$:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq \boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) \tag{1}$$

where $\boldsymbol{y}$ is a measured sinogram data and L denotes a negative Poisson log-likelihood function. The negative Poisson log-likelihood for emission tomography is defined as follows:

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) = \sum_i \bar{y}_i(\boldsymbol{f}) - y_i \log \bar{y}_i(\boldsymbol{f}) \tag{2}$$

where y_i is the ith element of the measurement $\boldsymbol{y}$ and

$$\bar{y}_i(\boldsymbol{f}) \triangleq [\boldsymbol{A}\boldsymbol{f}]_i + s_i$$

where $\boldsymbol{A}$ denotes the system model and s_i is a scatter component for the ith measurement.

For SPECT imaging, we can incorporate an attenuation map and a depth-dependent point spread function model including penetration tails [14] in the system matrix $\boldsymbol{A}$. In our simulation, we assumed known s_i, but in practice, this scatter component can be estimated by using a triple energy window (TEW) method or by Monte Carlo methods [15].

Unregularized image reconstruction in (1) is *ill-posed*. In this case, converged recontructed images are very noisy. There are usually three approaches to deal with this noise: First of all, one can stop iteration before convergence. However, more iteration may be necessary for recovering high-frequency information (*e.g.* details) of image. Secondly, one can use a post-reconstruction filter (*e.g.*, Gaussian filter) to reduce noise. Lastly, one can add a regularizer to (1) (*e.g.*, quadratic roughness penalty, non-local regularizer). When using non-local regularizers for 3D images, the computation-complexity is usually very high.

B. Non-local regularizer

Recently, NL regularizers have been shown to yield high-quality images in many image reconstruction problems [3]–[5], [16]. A NL regularizer R can be added to (1) as follows:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq \boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{f}) \tag{3}$$

where β is a regularization parameter and

$$\mathrm{R}(\boldsymbol{f}) \triangleq \sum_{i,j\in\Omega_i} p\left(\|\boldsymbol{N}_i\boldsymbol{f} - \boldsymbol{N}_j\boldsymbol{f}\|^2\right), \tag{4}$$

$\|\cdot\|$ is the L_2 norm, $\boldsymbol{N}_i$ is an operator on the image $\boldsymbol{f}$ such that $\boldsymbol{N}_i\boldsymbol{f}$ is a vector of image intensities that are on the cube-shaped patch around the ith voxel, and p is any potential function. A typical choice for the function p is [3], [16]

$$p(t) = \exp\left(-\frac{\|\boldsymbol{N}_i\tilde{\boldsymbol{f}} - \boldsymbol{N}_j\tilde{\boldsymbol{f}}\|^2}{2\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}\right)\frac{t}{2N_{\mathrm{f}}} \tag{5}$$

where $\tilde{\boldsymbol{f}}$ is an initial image from any analytical image reconstruction (*e.g.*, filtered back projection) [16] or an estimated image from the previous iteration $\boldsymbol{f}^{(n)}$ [3]. In this case, p depends on i and j. Yang *et al*. proposed to use a few non-convex potentials including the Welsh potential [17]

$$p(t) = \sigma_{\mathrm{f}}^2\left(1 - \exp\left(-\frac{t}{2\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}\right)\right). \tag{6}$$

Wang *et al*. proposed to use the Fair potential [18], [19]

$$p(t) = \sigma_{\mathrm{f}}^2\left(\sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}} + \log\left(1 + \sqrt{\frac{t}{\sigma_{\mathrm{f}}^2 N_{\mathrm{f}}}}\right)\right). \tag{7}$$

Note that both (6) and (7) do not depend on an initial image and (7) is convex while (6) is non-convex. It has been reported that non-convex functions yielded better image quality than a convex function [5].

We can also design NL regularizers that can incorporate high-resolution side information such as CT or MR images [7]–[9] for better image-quality. In this paper, we focus on the Fair potential in (7), but the proposed algorithm can be applied to any regularizer.

C. Alternating direction method of multipliers

We split the variable for the likelihood term and the regularizer term by replacing (3) with the following equivalent constrained optimization problem:

$$\hat{\boldsymbol{f}} \triangleq \underset{\boldsymbol{f}\geq \boldsymbol{0}, \boldsymbol{u}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{u}), \text{ sub. to } \boldsymbol{u} = \boldsymbol{f}. \tag{8}$$

By adding the augmented Lagrangian term, (8) becomes

$$\mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \beta \mathrm{R}(\boldsymbol{u}) + \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u} - \boldsymbol{d}\|^2 \tag{9}$$

where μ is a scalar value (design parameter) and $\boldsymbol{d}$ is a Lagrangian multiplier vector.

We can solve this optimization problem (9) by using the ADMM algorithm [20], [21] as follows:

For $n = 0, 1, 2, \cdots$

$$\boldsymbol{u}^{(n+1)} \in \underset{\boldsymbol{u}}{\operatorname{argmin}} \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}\|^2 + \beta \mathrm{R}(\boldsymbol{u}) \tag{10}$$

$$\boldsymbol{f}^{(n+1)} \in \underset{\boldsymbol{f}\geq \boldsymbol{0}}{\operatorname{argmin}} \mathrm{L}(\boldsymbol{y}|\boldsymbol{f}) + \frac{\mu}{2}\|\boldsymbol{f} - \boldsymbol{u}^{(n+1)} - \boldsymbol{d}^{(n)}\|^2 \tag{11}$$

$$\boldsymbol{d}^{(n+1)} = \boldsymbol{d}^{(n)} - (\boldsymbol{f}^{(n+1)} - \boldsymbol{u}^{(n+1)}) \tag{12}$$

End

where $\boldsymbol{f}^{(n)}$ is an estimated vector value $\boldsymbol{f}$ at the nth iteration. We can solve the sub-problems of (10) and (11) using existing methods.

We used the GD algorithm to solve (10) as follows:

$$\boldsymbol{u}^{(n+1)} = \boldsymbol{u}^{(n)} - \alpha \nabla \Phi^{(n)}(\boldsymbol{u}^{(n)}) \tag{13}$$

where α is a step size and

$$\Phi^{(n)}(\boldsymbol{u}) \triangleq \frac{\mu}{2}\|\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}\|^2 + \beta \mathrm{R}(\boldsymbol{u}).$$

We plug (13) into (10) to determine the step size as follows:

$$\alpha \in \underset{\alpha}{\operatorname{argmin}} \phi^{(n)}(\alpha) \tag{14}$$

where $\phi^{(n)}(\alpha) \triangleq \Phi^{(n)}(\boldsymbol{u}^{(n)} - \alpha \nabla \Phi^{(n)}(\boldsymbol{u}^{(n)}))$,

$$\nabla \Phi^{(n)}(\boldsymbol{u}) = \mu(\boldsymbol{u} - \boldsymbol{f}^{(n)} + \boldsymbol{d}^{(n)}) + \beta \nabla \mathrm{R}(\boldsymbol{u}),$$

$$\nabla \mathrm{R}(\boldsymbol{u}) = \sum_{i,j\in\Omega_i} (\boldsymbol{N}_i - \boldsymbol{N}_j)' 2\dot{p}\left(\|\boldsymbol{N}_i\boldsymbol{u} - \boldsymbol{N}_j\boldsymbol{u}\|^2\right)(\boldsymbol{N}_i - \boldsymbol{N}_j)\boldsymbol{u},$$

and $\dot{p}(t)$ is the first order derivative of $p(t)$. Since solving (14) is an intermediate step of solving (10), we do not need to find

an exact α value to minimize (14). We chose to use one step of Newton's method for (14) as follows [22]:

$$\alpha = -\frac{\dot{\phi}^{(n)}(0)}{\ddot{\phi}^{(n)}(0)} \tag{15}$$

where $\dot{\phi}^{(n)}(\alpha)$ and $\ddot{\phi}^{(n)}(\alpha)$ are the first and second order derivatives of $\phi^{(n)}(\alpha)$ with respect to α:

$$\dot{\phi}^{(n)}(0) = -\|\nabla\Phi^{(n)}(\boldsymbol{u}^{(n)})\|^2$$

and

$$\ddot{\phi}^{(n)}(0) \approx \nabla\Phi^{(n)}(\boldsymbol{u}^{(n)})' \left(\mu\nabla\Phi^{(n)} + \nabla \mathrm{R}(\nabla\Phi^{(n)})\right).$$

We approximate $\ddot{\phi}^{(n)}(\alpha)$ by not using the second derivative of $p(t)$ as suggested in [23, p. 683].

Equation (11) can be solved using De Pierro's EM algorithm [24] and OS approximation can be used to speed up the convergence rate. Whereas Wang *et al*. used De Pierro's algorithm with the surrogate function of their NL regularizer [4], we use De Pierro's algorithm with a shifted quadratic regularizer, which requires far less computation. One can find a similar formula for this sub-problem of (11) without a Lagrangian multiplier vector in [12].

III. Result

We simulated a 3D SPECT-CT system with the non-uniform attenuation map, collimator-detector response, and scatter component (128×21, $4.8^2\mathrm{mm}^2$ pixel size). We used the XCAT phantom [13] to generate the true SPECT image. The dimension of the SPECT image was $128 \times 128 \times 21$, $4.8^3\mathrm{mm}^3$ voxel size. We set the regularization parameters as follows: $\beta = 2^{-13}$, $\sigma_{\mathrm{f}} = 2^{1.5}$, the patch size $3 \times 3 \times 3$, the search neighborhood size $7 \times 7 \times 7$, five past estimated images for hessian approximation (L-BFGS-B), and $\mu = 2^{-7}$ (ADMM). Six subsets were used for OSEM and ADMM and 12 threads were used for computation (Intel Xeon 2.67GHz). A uniform initial image was used for all methods. We measured a normalized root mean square error (RMSE) for estimated images at all (outer) iterations and the definition of the RMSE is

$$\mathrm{RMSE} = \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}_{\mathrm{TRUE}}\|}{\|\boldsymbol{f}_{\mathrm{TRUE}}\|}. \tag{16}$$

Fig. 1 shows the plots of RMSE values versus computation time for different methods: GD, EM and OSEM using De Pierro's lemma, L-BFGS-B [25], and proposed ADMM. OSEM does not show much speed up as compared to EM due to computationally expensive NL regularizer calculation for all sub-iterations. ADMM separates the likelihood update and the regularizer update by splitting and runs more sub-iterations for the likelihood update (2 outer-iterations $\times 6$ subsets) than for the regularizer update (1 outer-iteration). These simulation results illustrate that repeated likelihood updates are more important for fast convergence than regularizer updates.

Fig. 2 shows estimated images of different methods at 500 seconds and the true image. At this early time, ADMM yielded the best contrast recovery among all other methods. Fig. 3 shows estimated images of different methods at 1000 seconds and the true image. In this case, L-BFGS-B achieved contrast recovery similar to ADMM. Both ADMM and L-BFGS-B yielded better contrast recovery than other methods.

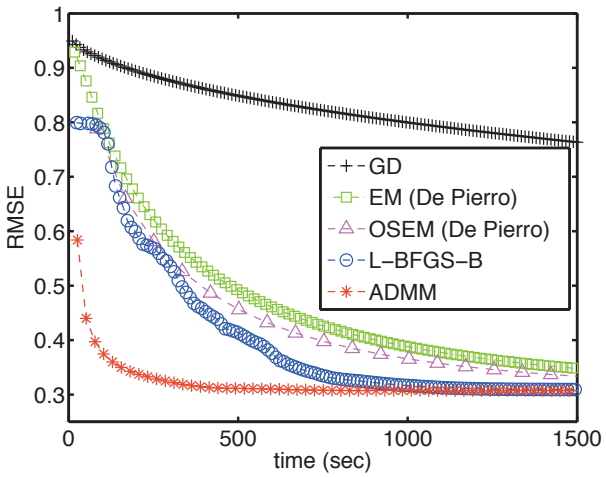


Fig. 1. RMSEs of estimated images using different algorithms over time. Proposed ADMM showed faster convergence rate than other previous methods.

IV. Discussion

We developed a new algorithm for emission tomography with computationally expensive NL regularizers using the ADMM. By combining with the OS approach, our proposed ADMM approached convergence much faster than existing methods such as GD, EM - De Pierro, OSEM - De Pierro, and L-BFGS-B. Since it seems more important to update the likelihood part frequently, our ADMM yielded faster convergence. Comparing our new method with other algorithms such as preconditioned conjugate gradient can be an interesting future work.

We demonstrated that our proposed method worked well for SPECT image reconstruction with the patch-based Fair potential function [4]. Our proposed method can be easily extended to other computationally expensive NL regularizers [3], [5], [16] and NL regularizers that use high-resolution side information [7]–[9] for both emission and transmission tomography.

ADMM requires a good μ value. Even though the theory of ADMM says that the algorithm will converge for any μ value, the choice of μ affects the convergence rate. Our future research will be an investigation on how to choose a good μ value.

Acknowledgment

The authors would like to thank Dr. Sathish Ramani for helpful discussion on ADMM.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[2] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–9, Dec. 1994.

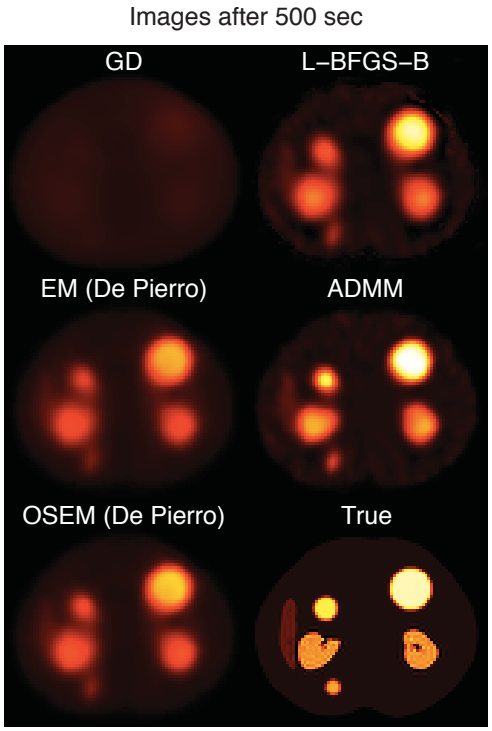


Fig. 2. Estimated images of different methods at 500 seconds and the true image. ADMM yielded the best contrast recovery among all other methods.

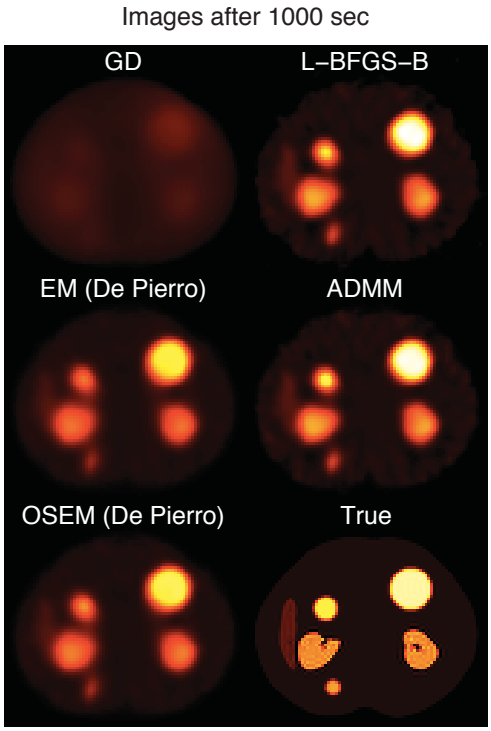


Fig. 3. Estimated images of different methods at 1000 seconds and the true image. L-BFGS-B achieved contrast recovery similar to ADMM. Both ADMM and L-BFGS-B yielded better contrast recovery than other methods.

[3] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 253–76, 2010.

[4] G. Wang and J. Qi, "Penalized likelihood pet image reconstruction using patch-based edge-preserving regularization," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 12, pp. 2194 –2204, dec. 2012.

[5] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: A unified variational framework," *IEEE Trans. Im. Proc.*, 2012.

[6] F. Rousseau, "A non-local approach for image super-resolution using intermodality priors," *Med. Im. Anal.*, vol. 14, no. 4, pp. 594–605, Aug. 2010.

[7] S. Y. Chun, J. A. Fessler, and Y. K. Dewaraja, "Non-local means methods using CT side information for I-131 SPECT image reconstruction," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2012, pp. 3362–6.

[8] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. Deroose, K. V. Laere, and J. Nuyts, "Evaluation of three mri-based anatomical priors for quantitative pet brain imaging," *IEEE Trans. Med. Imaging*, vol. 31, no. 3, pp. 599–612, 2012.

[9] V.-G. Nguyen and S.-J. Lee, "Anatomy-based PET image reconstruction using nonlocal regularization," in *Proc. SPIE 8313 Medical Imaging 2012: Phys. Med. Im.*, 2012.

[10] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 323–43, 2009.

[11] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Im. Proc.*, vol. 19, no. 12, pp. 3133–45, Dec. 2010.

[12] J. Xu, S. Chen, and B. M. W. Tsui, "Total variation penalized maximum-likelihood image reconstruction for a stationary small animal SPECT system," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011, pp. 225–8.

[13] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–8, Aug. 2008.

[14] S. Chun, J. Fessler, and Y. Dewaraja, "Correction for collimator-detector response in SPECT using point spread function template," *IEEE Trans. Med. Imaging*, vol. 32, no. 2, pp. 295–305, 2013.

[15] Y. K. Dewaraja, M. Ljungberg, and J. A. Fessler, "3-D Monte Carlo-based scatter compensation in quantitative I-131 SPECT reconstruction," *IEEE Trans. Nuc. Sci.*, vol. 53, no. 1, pp. 181–8, Feb. 2006.

[16] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *J. Sci. Comput.*, vol. 42, no. 2, pp. 185–197, Feb. 2010.

[17] M. Rivera and J. L. Marroquin, "Efficient half-quadratic regularization with granularity control," *Im. and Vision Computing*, vol. 21, no. 4, pp. 345–57, Apr. 2003.

[18] R. C. Fair, "On the robust estimation of econometric models," *Ann. Econ. Social Measurement*, vol. 2, pp. 667–77, Oct. 1974.

[19] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–46, Dec. 1990, corrections, T-MI, 10:2(288), June 1991.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[21] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1-3, pp. 293–318, Apr. 1992.

[22] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.

[23] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, 2nd ed. New York: Cambridge Univ. Press, 1992.

[24] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–7, Mar. 1995.

[25] J. L. Morales and J. Nocedal, "Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization"," *ACM Trans. Math. Softw.*, vol. 38, no. 1, pp. 7:1–7:4, Nov. 2011.

Simultaneous reconstruction of attenuation and activity in TOF-PET: analysis of the convergence of the MLACF algorithm.

Michel Defrise[1], Ahmadreza Rezaei[2], Johan Nuyts[2]

I. Introduction

Various algorithms have been recently proposed to estimate both the attenuation and the activity from time-of-flight (TOF) PET emission data [1-6], without requiring additional CT or MR information. Initial tests with simulated, phantom, and clinical data demonstrate that this simultaneous estimation problem can be solved with a surprisingly good accuracy. These results corroborate previous works demonstrating that the TOF emission data contain significant information on the attenuation [7,8], and they offer promising perspectives for various applications including for instance studies with partially known or mismatched CT data [5,6].

This work presents a mathematical and numerical analysis of the convergence of MLACF [4], a maximum likelihood algorithm that jointly estimates the activity distribution and the attenuation factors. This algorithm does not reconstruct the attenuation image, in contrast with the MLAA algorithm [3]. We consider the simplest case where there is no background due to scatter or random events, and demonstrate that i/ the MLACF algorithm is monotonic and asymptotically regular, ii/ if the likelihood function has a unique global maximum and if there are no local maxima, MLACF converges to the global maximizer. Although the mathematical analysis of the continuous version of the problem showed that noise free emission data determine the attenuation correction factors up to a scale factor [1,2,6], it is still unclear whether uniqueness also applies to the Poisson likelihood, except if the TOF data are consistent (uniqueness in that case was proven in [4]). Therefore, we present a numerical analysis of convergence on a simple 2D problem, the results of which suggests that local maxima of the likelihood, if any, would only occur for extremely low count data.

II. The MLACF algorithm

Let $\lambda_j \geq 0, j = 1, \ldots, M$ be the unknown activity image discretized on a grid of M voxels, and let $y_{i,t} \in I\!N_+, i = 1, \ldots, N, t = 1, \ldots, T$ be the measured emission data for line of response i and TOF bin t. We wish to find an activity estimate λ^* and an attenuation sinogram estimate a^* which maximize the log-likelihood

$$L(y, \lambda, a) = \sum_{i=1}^{N} \sum_{t=1}^{T} \{-a_i \, p_{i,t} + y_{i,t} \log(a_i \, p_{i,t})\} \qquad (1)$$

where $c_{i,j,t} \geq 0$ is the known system matrix, $0 < a_i \leq 1$ are unknown attenuation factors, and $p_{i,t} = \sum c_{i,j,t} \lambda_j$ is the expectation of the unattenuated data. Note the scale invariance $L(y, \lambda, a) = L(y, \alpha\lambda, a/\alpha)$ for any $\alpha > 0$, and the fact that L is concave in both λ and a, but not jointly concave in λ, a. Owing to the scale invariance, the constraint $a_i \leq 1$ can be ignored provided the unconstrained optimizer belongs to the open set

$$H = \left\{ x \in I\!R^M \,|\, x_j \geq 0, \text{ and } p_i = \sum_j c_{i,j} x_j > 0 \text{ if } y_i > 0 \right\} \qquad (2)$$

Thus ignoring the constraint, the likelihood is easily maximized w.r.t. a at fixed activity λ, yielding the ML estimate of the attenuation factor,

$$a_i^* = \frac{y_i}{p_i} \quad i = 1, \ldots, N \qquad (3)$$

where $y_i = \sum_t y_{i,t}$ and similarly $p_i = \sum_t p_{i,t}$. The same convention is used everywhere: quantities indexed by i denote quantities summed over the TOF bins. Substituting a^* in L reduces the problem to the maximization of the *reduced log-likelihood* (see [4] for details),

$$\tilde{L}(y, \lambda) = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t} \, \log \frac{p_{i,t}}{p_i}. \qquad (4)$$

The MLACF algorithm is based on optimization transfer [9], it maximizes at iteration n the following concave and separable surrogate of $\tilde{L}$:

$$\begin{aligned} \tilde{L}^{sur}(y, \lambda, \lambda^n) &= \sum_{i=1}^{N} \Big(- y_i \, \log p_i^n - \frac{y_i \, (p_i - p_i^n)}{p_i^n} \\ &+ \sum_{t=1}^{T} \sum_{j=1}^{M} \frac{y_{i,t} \, c_{i,j,t} \lambda_j^n}{p_{i,t}^n} \log(\frac{\lambda_j p_{i,t}^n}{\lambda_j^n}) \Big) \end{aligned} \qquad (5)$$

with $p_{i,t}^n = \sum_j c_{i,j,t} \lambda_j^n$ and $p_i^n = \sum_t p_{i,t}^n$, yielding the iteration [4]

$$\lambda_j^{n+1} = T(\lambda^n)_j = \frac{\lambda_j^n}{\sum_{i=1}^{N} \frac{y_i \, c_{i,j}}{p_i^n}} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{y_{i,t} c_{i,j,t}}{p_{i,t}^n}. \qquad (6)$$

[1]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, B-1090, Brussels, Belgium, mail: mdefrise@vub.ac.be, [2]Dept. of Nuclear Medicine, Katholieke Universiteit Leuven, B-3000, Leuven, Belgium, mail: ahmadreza.rezaei@uzleuven.be, johan.nuyts@uzleuven.be.

Like ML-EM, this algorithm maintains strict positivity if the initial activity estimate is strictly positive, this property also guarantees the absence of singularity because $p_{i,t}^n > 0$. However, convergence to zero cannot be excluded, i.e. one may have $\lambda_j^n \to 0$ for some voxels. Since $\tilde{L}$ is scale invariant, we also define the normalized MLACF iteration $T_{\mathcal{N}}(\lambda) = T(\lambda)/||T(\lambda)||$, with $||x||$ the euclidean norm in $\mathbb{R}^M$. See [4,5] for the extension of MLACF to the case with background, which will not be considered here.

Some notations will be needed. We define the active set of data bins for a given voxel as:

$$\tau_j = \{(i,t) \,|\, c_{i,j,t} > 0 \text{ and } y_{i,t} > 0\} \quad j = 1, \ldots, M \tag{7}$$

and adopt an equivalent definition for the TOF-summed data,

$$\iota_j = \{i \,|\, c_{i,j} > 0 \text{ and } y_i > 0\} \quad j = 1, \ldots, M. \tag{8}$$

Note that if $\tau_j = \emptyset$,

$$\frac{\partial \tilde{L}(y,\lambda)}{\partial \lambda_j} = \sum_{i \in \iota_j} \left\{ -\frac{y_i c_{i,j}}{p_i} \right\} \leq 0 \tag{9}$$

and therefore any maximizer of the reduced log-likelihood satisfies $\lambda_j = 0$ (because of the non-negativity constraint). If in addition $\iota_j = \emptyset$ the reduced log-likelihood does not depend on the value of voxel j, the maximizer is undefined and as a logical (but arbitrary) estimate we take $\lambda_j = 0$. These voxels without active data can be set to zero and must not be further considered when maximizing the reduced log-likelihood. Therefore we assume below that $\tau_j \neq \emptyset$ and $\iota_j \neq \emptyset$ for $j = 1, \ldots, M$. We also assume that $y_i \geq 1$ and $y_{i,t} \geq 1$ for all non-zero data bins. Finally we define the constants

$$\begin{aligned} \xi &= \min_j \min_{i \in \iota_j} c_{i,j} > 0 \quad , \eta = \min_j \min_{(i,t) \in \tau_j} c_{i,j,t} > 0 \\ \omega &= \max_i \sum_j c_{i,j} \quad , \quad \sigma = \max_{i,t} \sum_j c_{i,j,t} > 0. \end{aligned} \tag{10}$$

III. Properties of MLACF

Though previous tests with MLACF showed good convergence and encouraging practical results, two questions remain unanswered: i/ is the maximizer of the reduced likelihood unique, are local maxima or saddle points possible?, and ii/ does MLACF converge? This section investigates the latter question, while the former is adressed through numerical tests in section IV. Here we summarize our main results, including only brief comments on the proofs.

Lemma 1. The sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ is such that the sequence $\tilde{L}(y, \lambda^n), n = 0, 1, 2, \ldots$ is non-decreasing and converges.

• This is the standard monotonicity property of surrogate-based algorithms, and convergence uses the upper bound $\tilde{L}(\lambda) \leq 0$, which follows from $p_{i,t} \leq p_i$.

Lemma 2. Let $\tilde{\lambda} \in \mathbb{R}^M$ be any positive vector with $||\tilde{\lambda}|| = 1$.

$$\tilde{L}(y, T(\tilde{\lambda})) - \tilde{L}(y, \tilde{\lambda}) \geq C \frac{||T(\tilde{\lambda}) - \tilde{\lambda}||^2}{\max(1, ||T(\tilde{\lambda})||)} \tag{11}$$

with T the mapping (6) and $C = (1/2)\min(\xi/\omega, \eta/\sigma) > 0$.

• The proof follows the same line as in [10]. It is based on a 2-term Taylor development of the surrogate $\tilde{L}^{sur}(y, x, \tilde{\lambda})$ around its unique maximizer $T(\tilde{\lambda})$, and on a lower bound on the curvature of the surrogate (5).

Lemma 3. Asymptotic regularity. The sequence of normalized iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ is such that

$$\begin{aligned} &\lim_{n \to \infty} ||T(\lambda^n) - \lambda^n|| = \lim_{n \to \infty} ||T_{\mathcal{N}}(\lambda^n) - \lambda^n|| = 0 \\ &\lim_{n \to \infty} ||T(\lambda^n)|| = 1. \end{aligned} \tag{12}$$

• The proof follows easily from the previous Lemma. Note that asymptotic regularity is not sufficient to prove convergence of the algorithm.

Proposition 4. The sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ has an accumulation point λ^*. If $\lambda^* \in H$ (see eq. (2)), then $\nabla_j \tilde{L}(y, \lambda^*) = 0$ for any voxel satisfying $\lambda_j^* > 0$. All accumulation points of the sequence $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ have the same value of the reduced log-likelihood.

• It seems reasonable to assume that a limit point belongs to the set H because otherwise there are lines of response i for which events have been detected ($y_i > 0$) but which only receive contributions from voxels with zero activity ($p_i = 0$). This theorem does not guarantee convergence, and even if there is convergence, it may be to a local maximum or to a saddle point. Further investigation on the existence of saddle points and local maxima is therefore warranted. In favorable cases, one has the

Corollary 5. If the only stationary point of the reduced likelihood is a unique global maximum $\lambda^\dagger \in H$, then the sequence of *normalized* iterates $\lambda^{n+1} = T_{\mathcal{N}}(\lambda^n)$ with $\lambda_j^0 > 0$ converges to $\lambda^\dagger$.

IV. Numerical results

A. Simulation parameters

We digitized a 2D thorax phantom on a $M = 64 \times 64$ image with pixel size 8.027 mm. Simulated TOF-PET data were generated by forward projecting this phantom with radial pixel size 8.027 mm, 64 angular samples on $[0, \pi)$, and $T = 8$ TOF bins with sampling $\Delta\tau = 64.0$ mm. The TOF profile was a gaussian with FWHM 80 mm. The aim of this study is to get insight into convergence and uniqueness, hence this coarse discretization was chosen to allow performing a very large number of iterations with various initial estimates λ^0. All calculations were done in double precision, using a matched backprojector and without data subset. The emission and attenuation phantoms are shown in Figure 1. The phantom support is an ellipse with axes of 300 mm and 470 mm and the minimum attenuation factor a_i was 0.015. A vial with activity 0.5, diameter 40 mm, and water attenuation was added outside the phantom and used to scale the reconstructed activity image after reconstruction. Poisson noise was added to the data to generate three noisy data sets S1, S2 and S3 with respectively a total of 479705, 15990 and 3198 events, corresponding to

respectively 300, 10, and 2 average events in the maximum data bin $y_{i,t}$. A large number of data bins are equal to zero in data set S3 (see Figure 2), allowing to challenge the algorithm's behaviour at the edge of the admissible domain.

The MLACF iteration was run up to 10^5 iterations, starting with a uniform image estimate $\lambda_j^0 = 1$ and with a set of 20 random initial images generated as $\lambda_j^0 = 1 + 0.8R$ where R is a pseudo-random number with uniform distribution in $(0, 1)$.

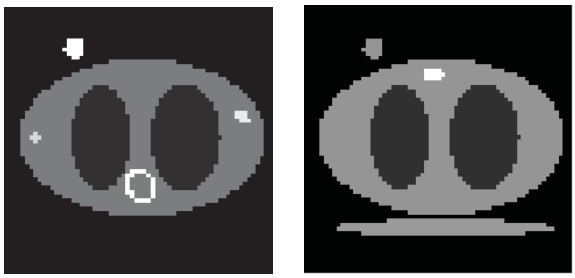


Fig. 1. The simulated phantom. Emission (left): activity is 0.2 (background tissues), 1.7 ("heart"), 0.05 ("lungs"), 0.40 and 0.45 ("tumors"), and 0.5 (vial). Attenuation (right): 0.00966/mm ((background tissues and vial), 0.00266/mm ("lungs"), 0.0187/mm ("spine"), and 0.01/mm ("bed").

Fig. 2. One central TOF-bin sinogram of the high noise data set S3, showing the large number of data bins for which no event have been detected.

B. Results

Figure 3 shows the convergence of the reduced log-likelihood cost function, with the three data sets and three of the random initial images. Note the irregular convergence, especially with data set S2. This phenomenon is tentatively attributed to the existence of regions where $\tilde{L}$ is almost flat for this specific data set. This behaviour is not observed when the same data sets are reconstructed with ML-EM assuming exact knowledge of the attenuation factors (Figure 4). Figure 5 shows numerical evidence for the asymptotic regularity of MLACF.

The three data sets have been reconstructed with MLACF, using a uniform initial activity and 20 random images. The relative difference between the maximum and minimum of these 21 values of the reduced likelihood after 10^5 iterations was $3.3\,10^{-14}$, $1.5\,10^{-9}$ and $1.77\,10^{-4}$ for S1, S2 and S3 respectively. The maximum relative RMSE difference between all pairs of images among the 21 reconstructions was $1.73\,10^{-8}$, $1.29\,10^{-7}$ and 0.25 for S1, S2 and S3 respectively. The difference between the two images corresponding to these maximum RMSE is negligible visually for S1 and S2. Even for the highest noise data set S3, the difference is small (Figure 6) and unlikely to be relevant in practice, where regularization would be applied. Nevertheless the plot of the RMSE between all pairs of images (Figure 7) reveals clusters; it is unclear at this point whether this observation reflects a lack of convergence, the digitization of the Poisson data (the number of events in the data bins of S3 take only 6 different values, between 0 and 5), or the convergence to different local maxima of the cost function. For the two other data sets, the corresponding RMSE values (plot not shown) are not only much smaller (see above) but do not show any structure or clustering.

Finally the loss of image quality caused by the absence of prior knowledge of the attenuation is illustrated for our example in Figure 8, which compares the MLACF and ML-EM reconstructions "at convergence" (meaning here 10^5 iterations).

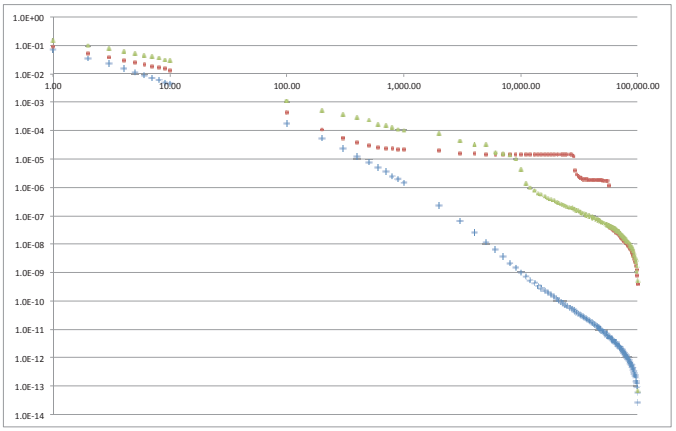


Fig. 3. The reduced log-likelihood $|\tilde{L}(y, \lambda^n) - \tilde{L}(y, \lambda^{100000})|/|\tilde{L}(y, \lambda^0)|$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles). The horizontal scale is the number of MLACF iterations n.

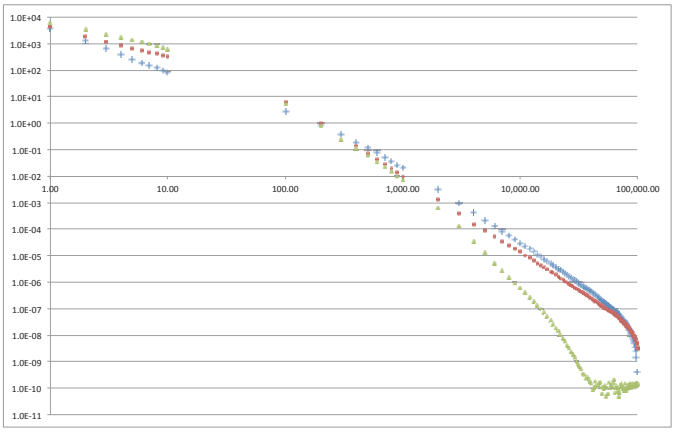


Fig. 4. The log-likelihood $|L(y, \lambda^n, a) - L(y, \lambda^{100000}, a)|$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles) reconstructed using ML-EM with the exact attenuation factors a_i. The horizontal scale is the number of ML-EM iterations n.

V. Conclusion

In practice the MLACF algorithm would be stopped before convergence and/or would be regularized by adding a smoothing penalty. Other modifications include an accelerated ordered-subset implementation and the generalization to account for a scatter or random background [4,5]. Understanding the convergence properties of the basic MLACF algorithm

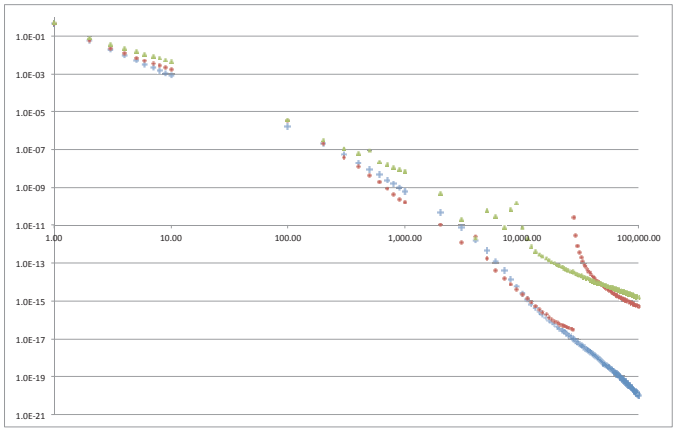


Fig. 5. The value of $||\lambda^{n+1} - \lambda^n||^2 / ||\lambda^n||^2$ for the data sets S1 (low noise, blue $+$), S2 (red circles), and S3 (high noise, green triangles). The horizontal scale is the number of MLACF iterations n.

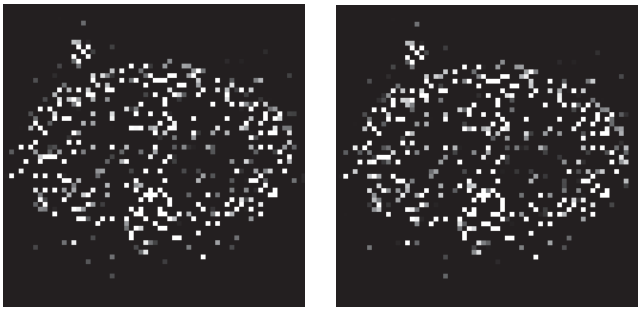


Fig. 6. The activity image reconstructed from data set S3 using the two initial images that yielded the largest difference. Grey scale $(0, 0.5)$.

(6) is nevertheless important. In this work we showed that MLACF is a monotonic and asymptotically regular algorithm, and that it converges to the maximum likelihood estimate of the activity image if the only stationary point of the reduced likelihood is a unique global maximizer, provided this maximizer satisfies condition (2). The results of the numerical study agree with these mathematical properties.

Uniqueness, up to the global scale factor, of the ML estimate is an open problem. The discrete Poisson model (1) investigated here does not include any specification on the range of the system matrix $c_{i,j,t} : \mathbb{R}^M \to \mathbb{R}^{N \times T}$. Clearly, uniqueness cannot hold in such a general setting, and counter-examples can easily be found, e.g. by building two disconnected voxel subsets that are not connected by any "active" ($y_{i,t} > 0$) line of response. In [2], the proof of uniqueness for the continuous model uses the range (consistency) conditions for the TOF Radon transform, and similar conditions will need to be added to the discrete model to give any hope of proving uniqueness. The numerical results seem to indicate that uniqueness does hold for the simple 2D problem considered, except possibly for very low statistics data such as S3.

References

[1] A. Rezaei et al. Simultaneous reconstruction of activity and attenuation in time-of-flight PET, IEEE Nucl Sci Symp Conf Record, Valencia, Oct 2011, MIC8-7.

[2] M Defrise et al. Time-of-flight PET data determine the attenuation sinogram up to a constant, Phys. Med. Biol., 2012, 57 (4), 885-899.

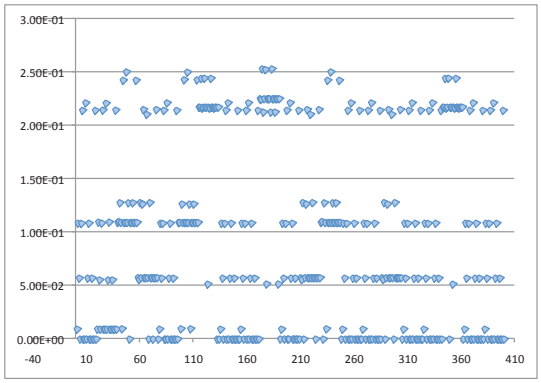


Fig. 7. Unordered plot of the relative RMSE differences between all pairs of images reconstructed from data set S3 with 100000 iterations of MLACF and various initial images. All points appear twice.

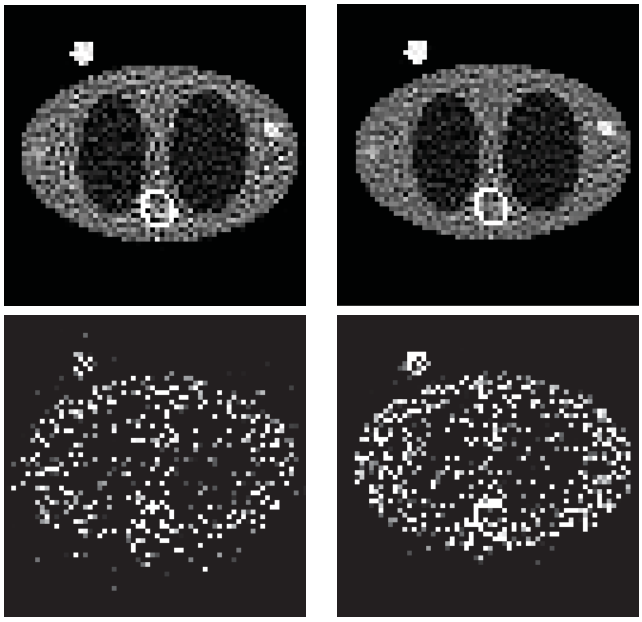


Fig. 8. The emission image reconstructed from data set S1 (top row) and S3 (bottom row) with 10^5 iterations. Left: MLACF. Right: ML-EM with known attenuation. Uniform initial images. Grey scale $(0, 0.5)$.

[3] A. Rezaei, et al, "Simultaneous reconstruction of activity and attenuation in time-of-flight PET", IEEE Trans. Med. Imag., 2012, 31 (12), 2224 - 2233.

[4] J. Nuyts et al., "ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M04-7.

[5] V. Panin et al."Reconstruction of Uniform Sensitivity Emission Image with Partially Known Axial Attenuation Information in PET-CT Scanners", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M04-1

[6] S. Ahn et al., "Convergent Iterative Algorithms for Joint Reconstruction of Activity and Attenuation from Time-of-Flight PET Data", IEEE Nucl Sci Symp Conf Record, Anaheim, Oct 2012, M19.5.

[7] M Conti. Why is TOF PET reconstruction a more robust method in the presence of inconsistent data, Phys. Med. Biol., 2011, 56, 155-168.

[8] A. Salomon et al. Simultaneous reconstruction of activity and attenuation for PET/MR, IEEE Trans Med Imaging, 2011, 30, 804-813.

[9] K. Lange et al., "Optimization transfer algorithms using surrogate objective functions", J. Comput. Graph. Stat., 2000, 9, 1-59.

[10] A. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography", IEEE Trans. Med. Imag., 1995, 14, 132-137.

Edge-Preserving PET Image Reconstruction Using Trust Optimization Transfer

Guobao Wang and Jinyi Qi

Abstract— **Iterative image reconstruction for positron emission tomography (PET) can improve image quality by using spatial regularization. The most commonly used quadratic penalty often over-smoothes edges and fine features in reconstructed images, while non-quadratic penalties can preserve edges and achieve higher contrast recovery. Existing optimization algorithms such as the expectation maximization (EM) and preconditioned conjugate gradient (PCG) algorithms work well for the quadratic penalty, but are less efficient for less-smooth or nonsmooth edge-preserving regularizations. This paper proposes a new algorithm to accelerate edge-preserving image reconstruction by using two strategies: trust surrogate and optimization transfer descent. Trust surrogate approximates the original penalty by a smoother function at each iteration, but guarantees the algorithm to converge monotonically; Optimization transfer descent accelerates a conventional optimization transfer algorithm by using conjugate gradient and line search. Results of computer simulations show that the proposed algorithm converges much faster than the conventional EM and PCG for smooth edge-preserving regularization and can also be more efficient than the current state-of-art algorithms for the nonsmooth ℓ_1 regularization.**

I. Introduction

Iterative image reconstruction methods can accurately model the system response and noise statistics in positron emission tomography (PET). They have been increasingly used to improve image quality [1]. Maximum likelihood (ML) method reconstructs image from projections by maximizing the log likelihood of PET data and can be elegantly solved by the expectation maximization (EM) algorithm [2]. However, a true maximum likelihood solution can be very noisy. Common ways to stabilize the image estimation are either terminating the iteration before convergence or using a penalty function to encourage spatially smooth images. Both early termination of the EM algorithm and using the quadratic penalty function tend to over-smooth edges and small objects in reconstructed images. Non-quadratic regularization can preserve edges and can achieve higher contrast recovery for small targets.

Existing optimization transfer (OT) algorithms (EM-based [2], [3] and others [4]) work well for quadratic regularization. However, when applied to a less-smooth nonquadratic penalty function, these algorithms can be very insufficient (as shown in the Simulation Studies Section) and they are not applicable to nonsmooth regularization because the penalty function is not differentiable at zero. While the preconditioned conjugate gradient (PCG) algorithm [5], [6] can be faster than OT algorithms, it still suffers the same problem of slow convergence because the widely used preconditioner was borrowed from the EM algorithm and contains no information of the regularization.

This work is supported by NIH grant R01EB00194.

G. Wang and J. Qi are with the Department of Biomedical Engineering, University of California, Davis, CA, USA.

There has been growing interests recently in developing algorithms for nonsmooth regularization due to the emerging area of compressive sensing. The alternating direction method of mulitpliers (ADMM) has been developed for Poissonian image deconvolution under different names (PIDAL [8] and PIDSplit [9]) and has been applied to PET image reconstruction [10]. ADMM-type algorithms can be very fast, but are not guaranteed to converge monotonically. The ADMM for Poisson data involves three parameters that have to be tuned for fast convergence, which is a nontrival task. SPIRAL [11] is another new algorithm that utilizes the fast iterative shrinkage and thresholding algorithm (FISTA). It is fast in its nonmonotone implementation, but can be slow if monotonicty is enforced.

In this paper, we develop a new algorithm that achieves fast monotonic convergence for edge-preserving image reconstruction. We first introduce the optimization transfer descent (OTD) concept by exploring the descent nature of the original optimization transfer for minimization problems. Conjugate gradient and line search [12, p.120] are then incorporated in the OTD for acceleration. The OTD can be viewed as a PCG algorithm with an implicitly defined preconditioner which contains information of both likelihood term and regularization term. It is therefore expected to achieve faster convergence than the conventional PCG algorithm. To extend the OTD algorithm to nonsmooth penalties, we adopt the trust surrogate concept that has been used in the Levenberg-Marquardt and trust region methods [12, p.262] for nonlinear optimization. We approximate the original objective function by a smooth surrogate and solve the smoothed surrogate by the OTD algorithm. The resulting trust optimization transfer algorithm can solve edge-preserving image reconstruction very efficiently. It is nearly free of parameter tuning and guarantees monotonic convergence.

II. Penalized Likelihood PET Reconstruction

A. PET Image Reconstruction

PET data $\boldsymbol{y} = \{y_i\}$ can be well modeled as a collection of independent Poisson random variables with the log likelihood function as

$$L\big(\boldsymbol{y}|\bar{\boldsymbol{y}}(\boldsymbol{x})\big) = \sum_{i=1}^{n_i} y_i \log \bar{y}_i(\boldsymbol{x}) - \bar{y}_i(\boldsymbol{x}). \tag{1}$$

The expected data $\bar{\boldsymbol{y}}(\boldsymbol{x})$ is related to the unknown image $\boldsymbol{x}$ through an affine transform

$$\bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{r} \tag{2}$$

where $\boldsymbol{P} = \{p_{ij}\} \in I\!R^{n_i \times n_j}$ is the system matrix with p_{ij} denoting the probability of detecting an event originated at

pixel j by detector pair i, $\boldsymbol{r}$ accounts for background events such as randoms and scatters. n_i is the total number of detector pairs and n_j is the total number of pixels in image.

Penalized likelihood (PL) reconstruction (or equivalently maximum *a posteriori*, MAP) estimates the unknown image by minimizing a penalized negative likelihood function

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\geq\boldsymbol{0}} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) = -L\big(\boldsymbol{y}|\bar{\boldsymbol{y}}(\boldsymbol{x})\big) + \beta U(\boldsymbol{x}) \quad (3)$$

where $U(\boldsymbol{x})$ is an image roughness penalty. Conventionally the image roughness is measured based on the intensity difference between neighboring pixels, either in an anisotropic form

$$U(\boldsymbol{x}) = \sum_{j=1}^{n_j} \sum_{k\in\mathcal{N}_j} w_{jk}\psi_\delta(x_j - x_k) \quad (4)$$

or in an isotropic form

$$U(\boldsymbol{x}) = \sum_{j=1}^{n_j} \psi_\delta\Big(\sqrt{\sum_{k\in\mathcal{N}_j} w_{jk}(x_j - x_k)^2}\Big) \quad (5)$$

where $\psi_\delta(t)$ is the penalty function and δ is a parameter that controls the smoothness of the penalty function, w_{jk} is the weighting factor related to the distance between pixel j and pixel k in the neighborhood $\mathcal{N}_j$, and β controls the strength of the regularization.

A penalty that can preserve edges is the nonsmooth ℓ_1 (absolute value function)

$$\psi_\delta(t) = |t| \quad (6)$$

which is not differentiable at zero. Smooth ℓ_1 functions that are differentiable at zero have also been proposed. One example is the Fair function

$$\psi_\delta(t) = \delta\big(\frac{|t|}{\delta} - \log(1 + \frac{|t|}{\delta})\big) \quad (7)$$

which has a continuous second-order derivative. Other examples of smooth ℓ_1 functions include the hyperbola function $\sqrt{t^2+\delta^2}$ and the Huber function. Smooth ℓ_1's approximate the quadratic function when $|t| \ll \delta$ and approach the ℓ_1 for $|t| \gg \delta$.

III. Optimization Transfer Descent

The basic idea of optimization transfer (a.k.a. majorization-minimization) is to construct a surrogate function $Q(\boldsymbol{x};\boldsymbol{x}^n)$ of the image $\boldsymbol{x}$ at the nth iteration which majorizes the original objective function $\Phi(\boldsymbol{x})$ by satisfying the following two conditions [7]:

$$Q(\boldsymbol{x};\boldsymbol{x}^n) - Q(\boldsymbol{x}^n;\boldsymbol{x}^n) \geq \Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}^n), \quad (8)$$

$$\boldsymbol{\nabla} Q(\boldsymbol{x}^n;\boldsymbol{x}^n) = \boldsymbol{\nabla}\Phi(\boldsymbol{x}^n). \quad (9)$$

where $\boldsymbol{\nabla}$ denotes the gradient with respect to $\boldsymbol{x}$. Then the minimization of $\Phi(\boldsymbol{x})$ is transferred into minimizing $Q(\boldsymbol{x};\boldsymbol{x}^n)$

$$\boldsymbol{x}_{\mathrm{OT}}^{n+1} = \arg\min_{\boldsymbol{x}\geq\boldsymbol{0}} Q(\boldsymbol{x};\boldsymbol{x}^n). \quad (10)$$

The surrogate function $Q(\boldsymbol{x};\boldsymbol{x}^n)$ is usually easier to optimize by design than the original objective function. The new update $\boldsymbol{x}_{\mathrm{OT}}^{n+1}$ decreases the original objective function monotonically,

$$\Phi(\boldsymbol{x}_{\mathrm{OT}}^{n+1}) \leq \Phi(\boldsymbol{x}^n). \quad (11)$$

By setting $\boldsymbol{x}^{n+1} = \boldsymbol{x}_{\mathrm{OT}}^{n+1}$, the majorization-minimization procedure guarantees monotonic convergence. The well-known expectation maximization (EM) algorithm [2] is a special case of the optimization transfer algorithms [7].

We observe from Eq. (11) that the OT direction

$$\boldsymbol{d}_{\mathrm{OT}}^{n+1} = \boldsymbol{x}_{\mathrm{OT}}^{n+1} - \boldsymbol{x}^n \quad (12)$$

is a descent direction. This inspires the following more aggressive update on $\boldsymbol{x}$:

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^n + \alpha\mathcal{C}\boldsymbol{d}_{\mathrm{OT}}^{n+1} \quad (13)$$

where $\mathcal{C}$ denotes the conjugacy operation that is used in nonlinear conjugate gradient (CG) algorithm, α is a step size determined by a line search:

$$\hat{\alpha} = \arg\min_{\alpha} \Phi(\boldsymbol{x}^n + \alpha\mathcal{C}\boldsymbol{d}_{\mathrm{OT}}^{n+1}). \quad (14)$$

A second line search [5] is used to enforce nonnegativity constraint on $\boldsymbol{x}$.

To distinguish the new update from the original OT update, we refer to it as the optimization transfer descent (OTD). The OTD algorithm moves more aggressively than the OT does, while still guaranteeing to converge monotonically.

IV. Trust Optimization Transfer

The OTD algorithm developed in the previous section is not directly applicable to the nonsmooth ℓ_1 penalty function. To extend the OTD algorithm to nonsmooth penalties, we borrow the trust surrogate concept from the classic Levenberg-Marquardt and trust region methods [12, p.262] for nonlinear optimization.

At iteration n, the original objective function $\Phi(\boldsymbol{x})$ is approximated by a surrogate function S:

$$S_\sigma(\boldsymbol{x};\boldsymbol{x}^n) = S^L(\boldsymbol{x};\boldsymbol{x}^n) + \beta S_\sigma^U(\boldsymbol{x};\boldsymbol{x}^n) \quad (15)$$

where $S^L(\boldsymbol{x};\boldsymbol{x}^n)$ is the surrogate of the likelihood term and $S_\sigma^U(\boldsymbol{x};\boldsymbol{x}^n)$ is the smooth approximation of the penalty term, with σ being the damping parameter. Note that $S_\sigma(\boldsymbol{x};\boldsymbol{x}^n)$ is not an optimization transfer surrogate and does not have to satisfy the two conditions in (8) and (9).

A new estimate is then obtained by minimizing the surrogate

$$\hat{\boldsymbol{\mu}}(\sigma) = \arg\min_{\boldsymbol{x}\geq\boldsymbol{0}} S_\sigma(\boldsymbol{x};\boldsymbol{x}^n), \quad (16)$$

which can be solved by the OTD algorithm. If

$$\Phi\big(\hat{\boldsymbol{\mu}}(\sigma)\big) \leq \Phi(\boldsymbol{x}^n), \quad (17)$$

then the associated surrogate $S_\sigma(\boldsymbol{x};\boldsymbol{x}^n)$ is so called trust surrogate. The image estimate is then updated by

$$\boldsymbol{x}^{n+1} = \hat{\boldsymbol{\mu}}(\sigma). \quad (18)$$

If Eq. (17) is not satisfied, a new value of σ will be tested until a trust surrogate is found.

To apply OTD, we use

$$S^L(\boldsymbol{x};\boldsymbol{x}^n) = -L\big(\boldsymbol{y}|\bar{\boldsymbol{y}}(\boldsymbol{x})\big) \quad (19)$$

$$S_\sigma^U(\boldsymbol{x};\boldsymbol{x}^n) = \sum_{j=1}^{n_j} \sum_{k\in\mathcal{N}_j} w_{jk}\psi_\sigma(x_j - x_k) \quad (20)$$

where σ is greater than or equal to δ ($\delta = 0$ for nonsmooth ℓ_1). For efficient computation, we do not solve the minimization in (16) completely. Instead, only one iteration of OTD is used and we check the monotonicity in $\Phi(\boldsymbol{x})$.

A. Search Rule

In order to determine σ at each iteration, we define ρ as the ratio between the changes by $\hat{\boldsymbol{\mu}}(\sigma)$ in the original cost function $\Phi(\boldsymbol{x})$ and in the surrogate function $S_\sigma(\boldsymbol{x}; \boldsymbol{x}^n)$:

$$\rho\big(\hat{\boldsymbol{\mu}}(\sigma)\big) = \frac{\Phi\big(\hat{\boldsymbol{\mu}}(\sigma)\big) - \Phi(\boldsymbol{x}^n)}{S_\sigma\big(\hat{\boldsymbol{\mu}}(\sigma); \boldsymbol{x}^n\big) - S_\sigma\big(\boldsymbol{x}^n; \boldsymbol{x}^n\big)}. \tag{21}$$

If $\rho\big(\hat{\boldsymbol{\mu}}(\sigma)\big) \geq 0$, then $\Phi\big(\hat{\boldsymbol{\mu}}(\sigma)\big) \leq \Phi(\boldsymbol{x}^n)$.

When $\rho\big(\hat{\boldsymbol{\mu}}(\sigma)\big)$ is large, the value of σ is trusted and will be used in next iteration. To prevent too many iterations being spent on the same value of σ with only a tiny decrease in $\Phi(\boldsymbol{x})$, we measure the relative change in $\Phi(\boldsymbol{x})$ by

$$\nu = \frac{\Phi(\boldsymbol{x}^{n+1}) - \Phi(\boldsymbol{x}^n)}{\Phi(\boldsymbol{x}^{n+1}) - \Phi(\boldsymbol{x}^{n_{\text{start}}(\sigma)})} \tag{22}$$

where $n_{\text{start}}(\sigma)$ denotes the index of the first iteration at which the current σ is used. The rule for determining σ is

$$\sigma = \begin{cases} \max(\delta, \sigma/\tau_1), & \rho \leq 0 \\ \sigma, & \rho > 0, \nu \geq \tau_3/\rho \\ \max(\delta, \sigma/\tau_2), & \rho > 0, \nu < \tau_3/\rho \end{cases} \tag{23}$$

where τ_1 and τ_2 are integers and τ_3 is a small percentage value. We use $\tau_1 = \tau_2 = 3, \tau_3 = 0.01$. If $\rho \leq 0$, σ should be decreased. If $\rho > 0$ and ν is greater than a threshold, the current value of σ will be used again. The threshold τ_3/ρ allows more iterations to be taken for a σ value that results in a large ρ. When ν is too small, the σ value shall be reduced in the next step even if $\rho > 0$.

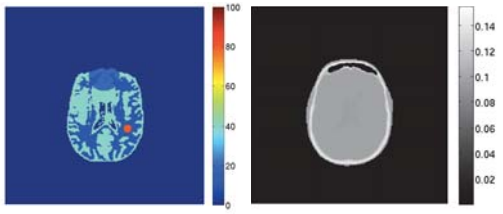
The initial value of σ is critical for the initial convergence speed. A large σ results in fast convergence for the surrogate optimization, while a σ closer to δ provides better approximation of the original objective function. We empirically find that $\sigma_0 = 0.01 \max(\boldsymbol{x})$ or $\sigma_0 = 0.1\text{mean}(\boldsymbol{x})$ is a good initial value if an estimate of $\boldsymbol{x}$ is known before reconstruction. An alternative is to determine σ_0 from the PET data by

$$\sigma_0 = 0.1\boldsymbol{p}^T(\boldsymbol{y} - \boldsymbol{r})/(\boldsymbol{p}^T\boldsymbol{p}),\ \boldsymbol{p} = \boldsymbol{P}\boldsymbol{1}. \tag{24}$$

V. Simulation Studies

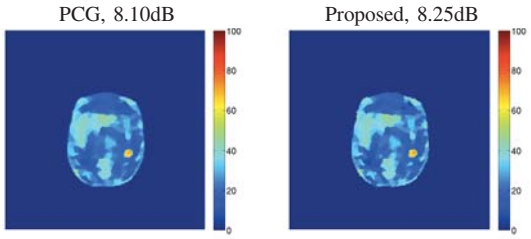
Computer simulation was conducted to compare proposed algorithm with several existing algorithms. We simulated a PET emission image using a 2D brain phantom (Fig. 1(a)). A real CT image was used to generate the attenuation factor (Fig. 1(b)). The PET image was first forward projected to generate a noise-free sinogram. A 20% uniform background was added to simulate mean randoms and scatters. Independent Poisson noise was then generated, resulting in a total of 200k events.

The noisy sinogram was reconstructed using two smooth ℓ_1s via the Fair function with $\delta = 10^{-2}$ and $\delta = 10^{-6}$, and the nonsmooth ℓ_1. The De Pierro's EM (DEM) [3] and PCG [5], [6] algorithms were used only for the smooth ℓ_1 regularizations, because they are not applicable to the

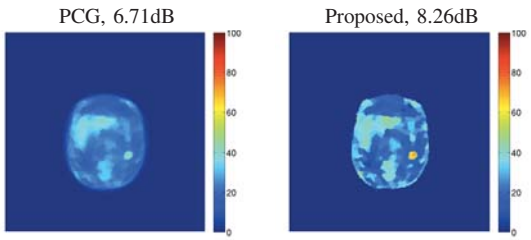


(a) Emission image (b) Attenuation map

Fig. 1. (a) The simulated PET emission image and (b) the attenuation map from a real CT image.



(a) $\delta = 10^{-2}$



(b) $\delta = 10^{-6}$

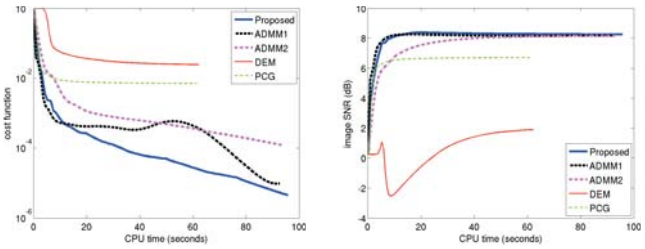
Fig. 2. PL reconstructions at 100 iteration using (a) the smooth ℓ_1 with $\delta = 10^{-2}$ and (b) smooth ℓ_1 with $\delta = 10^{-6}$ by the PCG and the proposed algorithms.

nonsmooth ℓ_1. The ADMM [8] with two different sets of three tuning parameters (ADMM1 and ADMM2) were implemented for both the smooth and nonsmooth penalties. ADMM1 was empirically tuned to converge as fast as possible in a non-monotonic fashion and ADMM2 was tuned to converge as fast as possible in a nearly monotonic way. The reconstruction step in the ADMM was solved by a PCG algorithm with 2 sub-iterations. The initial σ value in the proposed algorithm was calculated by Eq. (24). The SPIRAL algorithm [11], downloaded from the authors' website, was also used with its default parameter setting for the nonsmooth ℓ_1 penalty. The regularization parameter β was set to $\beta = 2^{-6}$ for the nonsmooth ℓ_1 and $\beta = 2^{-5}$ for the smooth ℓ_1's to achieve a good signal-to-noise ratio in the reconstructed images. All reconstructions start from the same uniform initial image.

To compare the convergence rate of different algorithms, we plotted the normalized objective function, which is defined as $E^n = (\Phi(\boldsymbol{x}^n) - \Phi^*)/(\Phi(\boldsymbol{x}^0) - \Phi^*)$ where Φ^* denotes a reference value of objective function, and the image signal-to-noise ratio (SNR), $SNR^n = -10\log_{10}\left(||\boldsymbol{x}^n - \boldsymbol{x}_{\text{true}}||^2/||\boldsymbol{x}_{\text{true}}||^2\right)$ as functions of CPU time.

A. Comparison for Smooth ℓ_1 Regularization

Fig. 2 shows the images reconstructed by the PCG and proposed algorithm using the smooth ℓ_1 penalty with two different δ values. As δ in the smooth ℓ_1 function decreases



(a) cost function (b) image SNR

Fig. 3. Convergence of (a) cost function and (b) SNR for the smooth ℓ_1 regularization with $\delta = 10^{-6}$ by different algorithms.

from 10^{-2} to 10^{-6}, conventional PCG becomes less efficient, resulting a substantial decrease in SNR. In comparison, the proposed algorithm is stable as δ changes, with a slight increase in image SNR for the smaller δ value.

Fig. 3 shows the cost function and image SNR as functions of CPU time for PL reconstruction using different algorithms. The smooth ℓ_1 penalty used $\delta = 10^{-6}$. Both the DEM and PCG have a slow convergence rate and result in a low image SNR. The two ADMM reconstructions converge very fast but the faster one behaves nonmonotonically. The proposed algorithm converges monotonically and is as fast as ADMM1.

B. Comparison for Nonsmooth ℓ_1

Different algorithms for PL reconstruction with nonsmooth ℓ_1 are compared. The reconstructed images at iteration 100 are shown in Fig. 4. The convergence plots of the cost function and image SNR are shown in Fig. 5 as a function of CPU time.

The SPIRAL converges fast and can achieve a very low cost function value but each iteration, especially the later iterations, takes much more CPU time than other algorithms. Both ADMM1 and SPIRAL have a nonmotonic behavior in the objective function value. While SPIRAL can be run in monotonic mode, it is slower than its nonmonotone implementation and may become extremely slow at later iterations.

The ADMM2 is slower than ADMM1 but converges almost monotonically and has a slightly higher SNR (8.31dB versus 8.29dB) at 100 iteration. Fig. 5 indicate that the proposed algorithm runs fastest among all algorithms in terms of minimizing the cost function and is as fast as ADMM1 in terms of SNR convergence.

VI. Conclusion

We have proposed a trust optimization transfer algorithm for edge-preserving PET image reconstruction. The fast convergence of the proposed algorithm is demonstrated using simulated data. Compared with the emerging algorithms such as the ADMM and SPIRAL algorithms for nonsmooth ℓ_1 regularization, the proposed algorithm is guaranteed to converge monotonically and its convergence speed is at least comparable to that of nonmonotoic ADMM and SPIRAL.

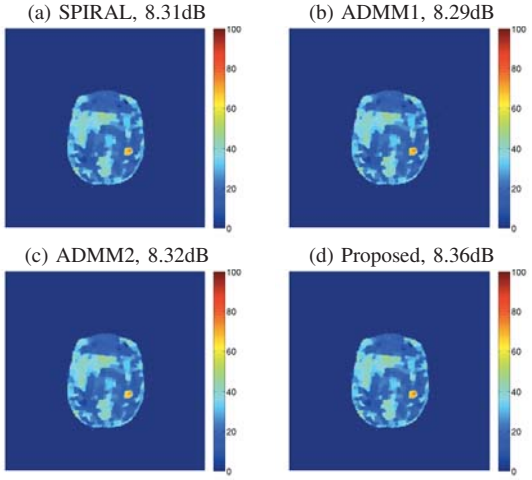


Fig. 4. PL reconstructions at 100 iteration using the nonsmooth ℓ_1 by SPIRAL, ADMMs and the proposed algorithm.

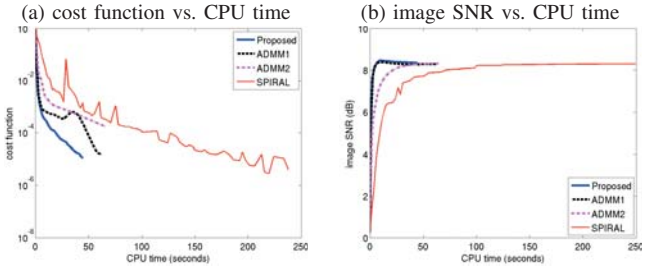


Fig. 5. Convergence of cost function and SNR as functions of CPU time for the nonsmooth ℓ_1 regularization by different algorithms.

References

[1] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Physics in Medicine and Biology*, vol. 51, no. 15, pp. R541-578, 2006.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[3] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1995.

[4] J. A. Fessler and H. Erdogan, "A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction," *1998 IEEE Nuclear Science Syposium and Medical Imaging Conference*, vol. 2, pp. 1132–5, 1998.

[5] L. Kaufman. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Transations on Medical Imaging*, 6(1):37–51, March 1987.

[6] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Y. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 687-701, 1994.

[7] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.

[8] M. A. T. Figueiredo, J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[9] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *Journal of Visual Communication and Image Representation*, vol. 21, no. 3, pp. 193–199, 2010.

[10] D. J. Lingenfelter, J. A. Fessier, "Augmented Lagrangian methods for penalized likelihood reconstruction in emission tomography," *2010 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 3288–3291 , 2010.

[11] Z. T. Harmany, R. F. Marcia, R. M. Willett, "This is SPIRAL-TAP: sparse Poisson intensity reconstruction aLgorithmstheory and practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

Combined AW-OSEM Reconstruction and Mass-Preserving Motion Correction of PET Data

Thomas Kösters*†, Michael Fieseler*, Tobias Block†, Frank Wübbeling‡, David Faul§, Fernando Boada†, Klaus Schäfers*

*European Institute for Molecular Imaging, University of Münster, Münster, Germany

†Department of Radiology, NYU Langone Medical Center, New York, New York

‡Department of Mathematics and Computer Science, University of Münster, Münster, Germany

§Siemens Medical Solutions, New York, USA

***Abstract*—Intrinsic motion is one of the major problems of quantitative PET. Several motion correction methods have already been described. In this work the combined reconstruction and motion correction approaches are analyzed. After studying the orientation of the motion fields, a possible way to speed up the algorithm is presented. Additionally, a mass-preserving scaling algorithm is introduced to account for loss and artificial creation of activity. Results on artifical data are presented.**

I. Introduction

Recently there have been lots of publications dealing with combined reconstruction and motion correction [1], [2], [3], [4]. It has been shown that these methods are superior compared to image based motion correction approaches. In this work we deal with problems which arise in the implementation of those algorithms. One topic is the correct choice of the direction of the motion fields. We introduce a mass-preserving scaling method which guarantees that no activity gets lost nor artifical activity is created during the reconstruction. Finally we show how to speed up already existing motion correction algorithms.

II. Methods

A. Attenuation Weighted OSEM

First we introduce our notation for attenuation weighted OSEM (AW-OSEM). The image is defined as a vector $b \in \mathbb{R}^V$ where V denotes the number of voxels. For L different lines of response the measured data is given by $g \in \mathbb{R}^L$. We divide g into three components, i.e., $g = g^T + g^S + g^R$, where g^T represents true coincidences, g^S scatter and g^R randoms or estimates of random coincidences. The system matrix is given by $X \in \mathbb{R}^{L\times V}$ where X represents the straight line model or the discretized X-ray transform, respectively. Hence we are solving $Xb = g$ for b and the OSEM algorithms reads

$$b_j^{\kappa+1} = \frac{b_j^\kappa}{\sum\limits_i x_{i,j}} \cdot \sum_i \left[\frac{x_{i,j} \cdot g_i}{\sum\limits_k x_{i,k} b_k^\kappa + g_i^R + g_i^S} \right],$$

where κ may indicate the current iteration or subset and j the corresponding voxel index. In order to correct for attenuation the elements of the system matrix are combined with the attenuation correction factors a_i, i.e., x_{ij} is replaced by x_{ij}/a_i, leading to

$$b_j^{\kappa+1} = \frac{b_j^\kappa}{\sum\limits_i (x_{i,j}/a_i)} \cdot \sum_i \left[\frac{x_{i,j} \cdot g_i}{\left(\sum\limits_k x_{i,k} b_k^\kappa\right) + a_i \cdot \left(g_i^R + g_i^S\right)} \right].$$

The attenuation correction factors using a corresponding linear attenuation map $\mu \in \mathbb{R}^V$ are given by

$$a_i = \exp\left(\sum_j x_{i,j} \mu_j \right).$$

This notation is similar to the notation for AW-OSEM used in [5]. For the sake of simplicity the normalization coefficients n_i are omitted. We mention that normalization may either be performed on the data before reconstruction or analogous to the just cited paper.

B. AW-OSEM with Motion Correction

As the next step we extend AW-OSEM for combined reconstruction and motion correction. We assume to work with $n = 1, \ldots, N$ gates which represent the different motion states. One of these N gates is selected as the reference gate; all other gates are called template gates. The aim of this approach is to reconstruct a single image using the acquired data of all gates to obtain a reconstruction of the reference gate with a higher signal-to-noise ratio. For each gate we have a set of sinograms of the same size as g. Hence g is extended from $\mathbb{R}^L$ to $\mathbb{R}^{N\cdot L}$. According to the larger dataset the system matrix has to be extended as well. We replace X by $P \in \mathbb{R}^{N\cdot L\times V}$. This new matrix can be divided into two components, i.e., $P = \mathbf{XM}$. The structure of $\mathbf{X}$ is rather simple since it is nothing else but a block matrix with X as defined above on the main diagonal

$$\mathbf{X} \in \mathbb{R}^{\mathbf{N\cdot L}\times\mathbf{N\cdot V}} = \begin{pmatrix} X \in \mathbb{R}^{L\times V} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & X \in \mathbb{R}^{L\times V} \end{pmatrix}.$$

Obviously, in case of $N = 1$ the matrix $\mathbf{X}$ is reduced to X. The second matrix M includes the motion information and is given by

$$M \in \mathbb{R}^{N\cdot V\times V} = \begin{pmatrix} \begin{pmatrix} m_{1,1} & \cdots & m_{1,V} \\ \vdots & \ddots & \vdots \\ m_{V,1} & \cdots & m_{V,V} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} m_{(N-1)\cdot V,1} & \cdots & m_{(N-1)\cdot V,V} \\ \vdots & \ddots & \vdots \\ m_{N\cdot V,1} & \cdots & m_{N\cdot V,V} \end{pmatrix} \end{pmatrix}.$$

Each column of each $V \times V$ sub-matrix represents the contributions of all voxels to a single voxel, i.e., column 1 includes the information which voxels are (partly) mapped to voxel 1. Each row represents the contribution of one single voxel to all

other voxels. More details on the interpretation of M will be discussed in the motion section II-C. Combining $\mathbf{X}$ and M, the elements of the new system matrix P can be written as

$$\begin{aligned} p_{i,j} &= \sum_{l=1}^{N\cdot V} \mathbf{x_{i,l}} \mathbf{m_{l,j}} \\ &= \sum_{l=1}^{N\cdot V} x_{i\%L,l\%V} m_{l,j} \end{aligned}$$

by using the modulo operator $\%$. Hence the components of the new system matrix are nothing else but a linear combination of the old matrix with weights according to the motion information. Therefore the AW-OSEM algorithm including motion correction reads

$$b_j^{\kappa+1} = \frac{b_j^{\kappa}}{\sum_i \left(p_{i,j}/a_i^m\right)} \cdot \sum_i \left[\frac{p_{i,j} \cdot g_i}{\left(\sum_k p_{i,k} b_k^{\kappa}\right) + a_i^m \cdot \left(g_i^R + g_i^S\right)} \right]$$

with a modified attenuation correction factor given by

$$a_i^m = \exp\left(\sum_j p_{i,j}\mu_j\right).$$

C. Motion Vectors

For this reconstruction approach it is necessary to know the motion of each single voxel in relation to the reference gate. Depending on the direction of the motion fields we know either where a voxel was moved to or where it originates from. We will see that both choices have their (dis-)advantages. Our suggestion is a combination of both approaches.

Before we discuss the implementation issues of the different methods we analyze the structure of M. In order to avoid a loss of activity / an artificial creation of activity during reconstruction, we postulate:

$$\sum_j m_{i,j} = 1, \tag{1}$$

i.e., one voxel must contribute 100% to other voxels. Although this restriction seems to be obvious it is not guaranteed in all cases. The next three sections discuss different ways of using the motion fields during reconstruction. The title of each subsection indicates the "direction" of the vectors used and may differ from the orientation of the given vector fields.

1) Template ⇒ Reference: As discussed in [3], this approach can be implemented straight-forward since it relies on interpolation only. During reconstruction of template data, each voxel is mapped to its "initial" position as illustrated in Figure 1. The advantage of this method is that classical [6] as well as more sophisticated algorithms [7] can be used to derive the system matrix which is joined with the interpolation step afterwards. In addition, equation 1 is satisfied per definition. Unfortunately, this method does not guarantee that each voxel of the reference gate will be hit when template data is used, because the interpolation needs not be surjective. Hence the interpolation may introduce errors leading to artifacts. Keywords for this method: *fast, error-prone*.

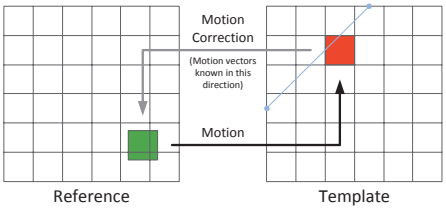


Fig. 1. For each voxel in the template image its "initial" position in the reference image is known. Hence each voxel can easily be moved during reconstruction. In this example linear interpolation is shown.

2) Reference ⇒ Template: In this case the exact movement for each voxel of the reference image into the template image is known. During reconstruction of template data the probability of particle emission in the red voxel and their measurement on the corresponding line of response has to be calculated. The image update must be performed at the green voxel. Since the red voxel is not on the grid of the reference image the computation of the system matrix is very time demanding [4]. Lamare calls this method exact, however condition (1) usually will not be fulfilled. In the worst case all voxels of the reference gate point to a single voxel in the template image. Hence we propose to introduce an appropriate scaling of the matrix elements $m_{i,j}$ which relates all vectors pointing on the same voxels to satisfy condition (1). In contrast to the method presented in section II-C1, this approach guarantees surjectivity which prohibits errors in the reconstruction if used in combination with the described scaling. Keywords for this method: *slow, exact*.

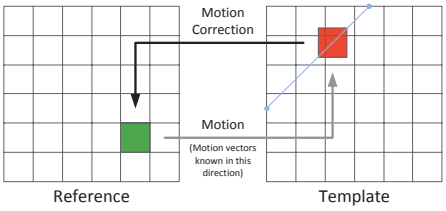


Fig. 2. For each voxel in the reference image the position after motion is known. Again linear interpolation is shown.

3) Template ⇒ Reference (Inverted Motion Fields): This third method combines the advantages of the other two methods to obtain an exact and fast algorithm. The first method is fast because of the simple interpolation step combined with the standard reconstruction framework. Hence our new method should work in a similar fashion to be likewise fast. To join this feature with the exact calculations of the second method we introduce a preprocessing step. Instead of calculating each $m_{i,j}$ during the reconstruction, the fraction of each reference image voxel on each template image voxel is precalculated and stored in a lookup table. As shown in Figure 3, the overlap of the red voxel on each yellow voxel is calculated. During the reconstruction we may now work with the yellow voxels and hence use the interpolation strategy as in section II-C1. The precalculation step can be interpreted as an inversion of the motion field and results in a surjective interpolation scheme. In

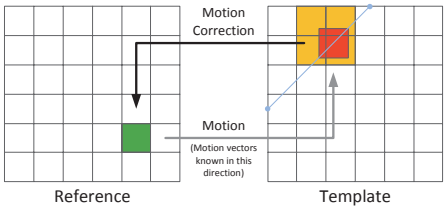


Fig. 3. The same vectors as in section II-C2 are used. In order to reconstruct using a similar approach as in section II-C1, the vectors are inverted.

order to satisfy the mass preserving condition (1), the values in the lookup table are scaled accordingly. Keywords for this method: *fast, exact*.

We mention that similar to the problem described in section II-C1 voxels in the template gate may not be hit. However, we did not experience any problems with this fact so far. Nevertheless, this problem should be analyzed for completeness.

D. Data

In order to evaluate the different motion correction approaches the XCAT phantom [8] was used. This enables us to use ground-truth data as well as ground-truth motion vectors. Hence the results do not depend on the estimation of the motion fields.

In this study we focused on respiratory motion. Eight different respiratory motion states were simulated using forward-projections and additional poisson noise. For comparison using a similar signal-to-noise ratio, eight noise realizations of the reference gate were simulated, summed up and averaged. Before adding the noise, the negative attenuation coefficients are used to include the effect of attenuation.

It is well-known [9] that the image based methods are working well in case each single gate has a reasonable signal-to-noise ratio. Therefore we are interested to compare the motion correction approaches when the single gates do no longer contain meaningful information.

The scanner geometry used for the numerical examples refers to the Siemens Biograph Sensation 16. The reconstructed volume has the size $175 \times 175 \times 47$ and the corresponding michelogram $192 \times 192 \times 175$. So far only linear interpolation was used to apply the motion correction in image space and data space.

E. Reconstruction

We compare the runtimes and results of several reconstruction and motion correction approaches, respectively. For all reconstructions we choose four iterations and five subsets. All times were measured on a shared-memory system with 48 GB RAM and four Intel(R) Xeon(R) CPU E7430 each consisting of four cores. The runtimes were measured using eight of the sixteen possible cores. The different reconstruction methods were implemented in the EMrecon reconstruction framework [10] and will be described briefly in the next sections. The method presented in section II-C2 is not investigated in this work since the implementation can be arbitrarily slow.

1) Reconstruction without Motion Correction (One Gate): This method represents the classical AW-OSEM reconstruction of a single gate as presented in section II-A. In order to obtain a motion contaminated dataset, the eight gates representing the different motion states were summed up and scaled according to the number of gates.

2) Reconstruction without Motion Correction (All Gates): Since we want to compare our relative reconstruction times to the results presented by Lamare [4] who used Listmode, we performed this additional reconstruction. It is required to reconstruct from the identical dataset (same amount of LORs) as used in the motion corrected reconstruction for a fair comparison. Regarding the system matrix $P = \mathbf{XM}$, this case refers to

$$\begin{aligned} m_{(n-1)\cdot V+j,j} &= 1 \quad n = 1,\ldots,N, j = 1,\ldots,V \\ m_{i,j} &= 0 \quad \textbf{otherwise}. \end{aligned}$$

The resulting image is identical compared to the result of section II-E1.

3) Reconstruction with Motion Correction (Template $\Rightarrow$ Reference): Here we are using the method presented in section II-C1. Since only the ground-truth vectors pointing in the opposite direction were available, the Matlab function TriScatteredInterp was used to obtain vectors pointing from the template gates to the reference. The resulting motion field obviously is only an approximation of the ground-truth data. Hence we only look at the runtime for comparison but not at the reconstruction result itself. This is reasonable because it already has been shown [4] that this method is inferior to the other methods.

4) Reconstruction with Motion Correction (Template $\Rightarrow$ Reference, Inverted Motion Fields): The given motion fields are used as presented in section II-C3. Therefore the inversion step is not performed during reconstruction but as a preprocessing before reconstruction.

5) Reconstruction with Motion Correction (Template $\Rightarrow$ Reference, Inverted Motion Fields, Mass-Preserving): Along the lines of section II-E4 the motion fields are preprocessed and used in the reconstruction. In addition, the mass-preserving scaling is introduced to satisfy equation (1).

III. Results and Discussion

The runtimes for the different reconstruction algorithms can be found in table I. The runtime of algorithm (II-E3) is about 10% longer compared to the uncorrected reconstruction given by (II-E2). This is comparable to the 5% given by Lamare [4]. The new proposed algorithms II-E4, II-E5 take the same time as algorithm (II-E3). Compared to the 10 times

Reconstruction Type	Time in s pro subset	Time in s total	Time relative
II-E1	1.44	28	0.19
II-E2	7.55	151	1
II-E3	8.32	166.33	1.1
II-E4	8.27	165.33	1.09
II-E5	8.3	166	1.1

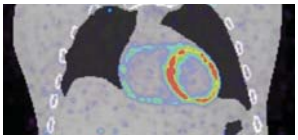
TABLE I
Reconstruction times for respiratory motion correction.

longer described by Lamare, this is a huge speedup. One could argue that the reconstruction may now be faster because the preprocessing takes so long. Regarding table II we can see that the preprocessing can easily be done and hardly takes extra time. In addition, the filesize of the stored motion data is rather small.

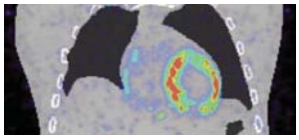
Reconstruction Type	Time in s for preprocessing	Filesize in MB
II-E3	7.3	240
II-E4	8	238
II-E5	8	238

TABLE II
PREPROCESSING TIMES AND FILESIZE OF MOTION DATA FOR RESPIRATORY MOTION CORRECTION.

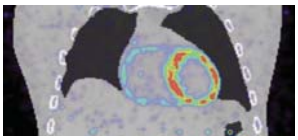
The improvement introduced by the mass-preserved scaling can be seen in Fig. 4. Although the reconstruction without mass-preserving seems to deliver reasonable results (see Fig. 4(c)), the scaling enables a higher recovery of activity inside the ventricle as can be seen in Fig. 4(d) and Fig. 4(e).



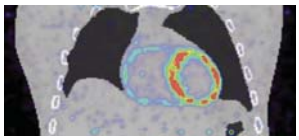
(a) Reconstruction of eight noise realizations of the reference gate.



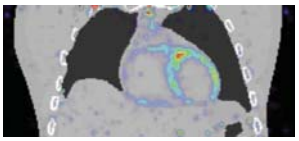
(b) Uncorrected reconstruction of motion contaminated data. The corresponding sinograms of all motion states are reconstructed as described in section II-E2.



(c) Reconstruction using the data based *non mass-preserving motion* correction approach as presented in section II-E4.



(d) Reconstruction using the data based *mass-preserving* motion correction approach as discussed in section II-E5.



(e) Difference of both data based motion correction approaches, i.e. Fig. 4(d) - Fig. 4(c). For better visual comparison the colormap has been scaled to a different level.

Fig. 4. Results for motion correction in case of respiratory motion. The data based approaches shown in 4(c),4(d) are able to produce reconstruction results close to the reference reconstruction shown in 4(a). The newly introduced mass-preservation scaling improves the recovery of update in the ventricle.

IV. CONCLUSION

We presented a possible way to speed up existing motion correction reconstruction algorithms which lead to a remarkable speedup. In case of datasets which include more motion than presented in this work, the reconstruction time compared to the uncorrected reconstructions may increase. Since this affects all discussed algorithms, the speedup will not get lost.

Additionally, we added the mass-preserving condition to guarantee an exact algorithm. We mention that our definition of mass-preserving motion correction differs from existing approaches in the literature. Gigengack et al. [11] and Dawood et al. [12] derive mass-preserving vectors but this does not replace the necessity of our mass-preserving approach.

In combination, the two presented modifications seem to be a good package to tackle the problem of motion correction in case of bad statistic data.

The next step will be the application of these algorithms to real patient data. Here it must be possible to obtain the motion vectors even in case of low PET signal. Therefore these kind of algorithms should suit optimally to combined MR/PET scanners like the Siemens Biography mMR. First datasets are under investigation.

REFERENCES

[1] I. Polycarpou, C. Tsoumpas, and P. K. Marsden, "Analysis and comparison of two methods for motion correction in PET imaging," *Medical Physics*, vol. 39, no. 10, pp. 6474–6483, 2012.

[2] S. Y. Chun, T. G. Reese, J. Ouyang, B. Guerin, C. Catana, X. Zhu, N. M. Alpert, and G. E. Fakhri, "MRI-Based Nonrigid Motion Correction in Simultaneous PET/MRI," *Journal of Nuclear Medicine*, vol. 53, no. 8, pp. 1284–1291, 2012.

[3] F. Qiao, T. Pan, J. W. C. Jr, and O. R. Mawlawi, "A motion-incorporated reconstruction method for gated PET studies," *Physics in Medicine & Biology*, vol. 51, p. 37693783, 2006.

[4] F. Lamare, M. J. L. Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C. C. L. Rest, A. J. Reader, and D. Visvikis, "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations," *Physics in Medicine & Biology*, vol. 52, pp. 5187–5204, 2007.

[5] M. E. Casey, "Point Spread Function Reconstruction in PET," Siemens, Tech. Rep., 2007.

[6] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Medical Physics*, vol. 12, pp. 252–255, 1985.

[7] D. J. Kadrmas, "Rotate-and-Slant Projector for Fast LOR-Based Fully-3-D Iterative PET Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 27, no. 8, pp. 1071–1083, August 2008.

[8] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, pp. 3800–3808, 2008.

[9] M. Dawood, F. Buether, X. Jiang, and K. Schaefers, "Motion Correction on 3D PET/CT Images Using Advanced Optical Flow Algorithms," *IEEE Transactions on Medical Imaging*, vol. 27, no. 8, pp. 1164–1175, 2008.

[10] T. Koesters, K. P. Schaefers, and F. Wuebbeling, "EMrecon: An expectation maximization based image reconstruction framework for emission tomography data," *IEEE NSS/MIC Conference Record*, pp. 4365–4368, 2011.

[11] F. Gigengack, L. Ruthotto, M. Burger, C. H. Wolters, X. Jiang, and K. P. Schaefers, "Motion Correction in Dual Gated Cardiac PET Using Mass-Preserving Image Registration," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 698–712, 2012.

[12] M. Dawood, F. Gigengack, X. Jiang, and K. P. S, "A mass conservation-based optical flow method for cardiac motion correction in 3D-PET," *Medical Physics*, vol. 40, 2013.

Quadratic Regularization Design for 3D Axial CT

Jang Hwan Cho and Jeffrey A. Fessler

***Abstract*—While iterative reconstruction (IR) methods have potential advantages over conventional FBP reconstruction such as reduced patient dose and improved noise properties, their use of statistical weighting and space variant scanning geometries can lead to nonuniform and anisotropic spatial resolution. Due to the large number of voxels in the image volume, regularization design methods based on discrete Fourier transforms would require prohivitive computational cost. In this paper, we propose a quadratic regularization design method for 3D axial X-ray computed tomography (CT) that aims to improve resolution isotropy and uniformity. Simulations and a phantom experiment show that the proposed method leads to more uniform and isotropic spatial resolution in 3D axial CT with modest computational cost.**

I. Introduction

Improved noise and spatial resolution properties are one of the potential advantages of statistical image reconstruction methods over conventional filtered back-projection (FBP) reconstruction [1]. Regularized image reconstruction methods, such as penalized weighted least squares (PWLS) method or a penalized-likelihood (PL) method, provide noise control by integrating a roughness penalty into the cost function. Although statistical weighting and system models are responsible for improving image quality, their interaction with a conventional quadratic roughness penalty results in images as anisotropic and nonuniform spatial resolution. This holds even for idealized shift-invariant imaging systems [2], and becomes most severe near the end slices of 3D axial or helical CT.

Several previous regularization designs aim to match the local impulse response of the estimator to a target impulse response by matrix manipulations and discrete Fourer transforms [2], [3]. The matrix and FFT methods need too much computation when applied to an entire image volume. Expecially for 3D axial or helical CT. A fast analytical regularization design method for 2D fan-beam X-ray CT that uses continuous space analogs to simplify the regularization design problem was proposed in [4]. In [5], the authors addressed the problem for 3D axial CT, but for a simplified 3D system that was modeled as a stack of 2D fan-beam systems. In this paper, we propose a regularization design for 3D axial X-ray computed tomography (CT) accounting for cone angle. Simulations and a phantom experiment show that the proposed method leads to more uniform and isotropic spatial resolution in 3D axial CT with modest computational cost.

This work was supported in part by NIH grant R01-HL-098686 and by equipment donations from Intel. The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA. Email: janghcho@umich.edu, fessler@umich.edu.

II. Method

A. Local Impulse Response

Consider a penalized weighted least squares (PWLS) objective function of the form

$$\Psi(\boldsymbol{x}) = \mathsf{L}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}), \;\; \mathsf{L}(\boldsymbol{x}) = \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_{\boldsymbol{W}}^2, \quad (1)$$

where $\boldsymbol{y}$ is the measurement vector, $\boldsymbol{A}$ is the system matrix, $\boldsymbol{x} = (x_1, \cdots, x_N)$ is the discretized version of the object being imaged, and $\boldsymbol{W} = \mathrm{diag}\{w_i\}$ is a statistical weighting matrix. A conventional quadratic regularizer is expressed as

$$\mathsf{R}(\boldsymbol{x}) = \beta \sum_j \sum_{l=1}^{N_l} \kappa_{l(j)} \kappa_j r_j^l \frac{1}{2} \left((c_l ** * x)[n,m,z] \right)^2, \quad (2)$$

where index j is a lexicographical ordering of $[n,m,z]$, N_l is the number of neighbors (13 in 3D), c_l is a function performs finite differences between neighboring voxels (see (11) below), κ's are the user-defined weights [2] for controlling spatial resolution in the reconstructed image, and $\{r_j^l\}$ are the directional regularizer coefficients that we will design.

For a quadratic regularizer, the local impulse response (LIR) at the jth voxel for the PWLS estimator is given as:

$$l^j = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}]^{-1} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j, \quad (3)$$

where $\boldsymbol{R}$ is the Hessian of the regularizer $\mathsf{R}(\boldsymbol{x})$ and $\boldsymbol{\delta}_j$ denotes an impulse function at jth voxel [2]. Our purpose is to design $\boldsymbol{R}$ such that our local impulse response l^j matches a target l^o that has more isotropic spatial resolution, at every pixel j. We simplify this process by turning to the frequency domain.

Assuming $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j$ and $\boldsymbol{R}\boldsymbol{\delta}_j$ are approximately locally circulant [6], we can approximate (3) as follows:

$$L^j = \frac{F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)}{F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right) + \beta F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)}, \quad (4)$$

where $F(\cdot)$ denotes the 3-D DFT.

Instead of directly using the discrete Fourier transform, we use the continuous-space analog of $H_j \triangleq F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)$ in spherical coordinates $\boldsymbol{\nu} \triangleq (\rho, \Phi, \Theta)$. We use a closed-form approximation for H_j that was suggested in [7]:

$$\begin{aligned} H_j(\boldsymbol{\nu}) &\approx K J(\boldsymbol{\nu}) \frac{\tilde{w}_j(\Phi)}{\rho \cos(\Theta)} \quad (5) \\ K &= \Pi \Delta_x^3 \Delta_z D_{sd}^2 / D_{so}^2 \\ J(\boldsymbol{\nu}) &= \mathrm{sinc}(\Delta_x \rho \cos(\Theta)\cos\Phi)^2 \\ &\quad \times \mathrm{sinc}(\Delta_y \rho \cos(\Theta)\sin\Phi)^2 \mathrm{sinc}(\Delta_z \rho \sin(\Theta))^2 \\ \tilde{w}_j(\Phi) &= \sum_{\beta \in B_j(\Phi)} \frac{\bar{w}_{\beta,j}}{d_{\beta,j} \sqrt{1 - (\zeta^j \cos(\theta^j))^2 \cos^2(\phi^j - \Phi)}}, \end{aligned}$$

where $D_{so} \cdot (\zeta^j, \phi^j, \theta^j)$ denotes the location of the jth voxel in spherical coordinates, K is a constant depending on voxel

sizes and scanner geometry, $J(\boldsymbol{\nu})$ is a factor depending only on spatial frequencies, $\bar{w}_{\beta,j} \triangleq w_\beta(\bar{s}_j^*)$ where $\bar{s}_j^*$ is the positon on the detector that maximizes the footprint of voxel j at source angle β, $d_{\beta,j}$ is the distance from the source to the xy-projection of voxel j, and $B_j(\Phi)$ is the set of the values of β for which the ray passing through voxel j is perpendicular to the frequency vector $\boldsymbol{\nu}$ where the ray and frequency vector are both projected onto the xy-plane [7]. Substituting (6) into (4) yields the following expression for the continuous space analog of L^j:

$$L^j \approx \frac{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta)) + \beta R_j(\boldsymbol{\nu})}, \tag{6}$$

where $R_j(\boldsymbol{\nu})$ is the local frequency response for the regularizer near pixel j (see (15) below).

B. Target Impulse Response

The local frequency response associated with penalized unweighted reconstruction is isotropic at the isocenter for a full scan, so we use it as our target response. At isocenter, (6) for uniform weights $(w_i = 1)$ is given as

$$\begin{aligned} H_o(\boldsymbol{\nu}) &\approx KJ(\boldsymbol{\nu})\frac{\tilde{u}_o(\Phi)}{\rho\cos(\Theta)}, \\ \tilde{u}_o(\Phi) &= |B_j(\Phi)|. \end{aligned} \tag{7}$$

Now the target local frequency response is

$$L^o \approx \frac{KJ(\boldsymbol{\nu})\tilde{u}_o(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{u}_o(\Phi)/(\rho\cos(\Theta)) + \beta R_o(\boldsymbol{\nu})}, \tag{8}$$

where L^o is the continuous-space analog of L^o.

Our purpose is to match the local impulse response at jth voxel to the target impulse resonse, i.e., we want

$$\begin{aligned} L^j &\approx \frac{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta)) + \beta R_j(\boldsymbol{\nu})} \\ &\approx \frac{KJ(\boldsymbol{\nu})\tilde{u}_o(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{u}_o(\Phi)/(\rho\cos(\Theta)) + \beta R_o(\boldsymbol{\nu})} \approx L^0. \end{aligned} \tag{9}$$

Cross multiplying and simplifying yields the goal

$$\tilde{u}_o(\Phi)R_j(\boldsymbol{\nu}) \approx \tilde{w}_j(\Phi)R_o(\boldsymbol{\nu}). \tag{10}$$

C. Regularization Structure

We first define a first-order differencing function that penalizes lth neighbor as

$$c_l = \frac{1}{\sqrt{n_l^2 + m_l^2 + z_l^2}}(\delta(n,m,z) - \delta(n-n_l, m-m_l, z-z_l)), \tag{11}$$

where n_l, m_l, z_l denote the offset of the neighbor. Taking the Fourier transform of (11) yields the following expression for the local frequency response $|C_l(\omega_1,\omega_2,\omega_3)|^2$

$$\begin{aligned} &= \frac{1}{n_l^2 + m_l^2 + z_l^2}\left|1 - e^{-i(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l)}\right|^2 \\ &= \frac{1}{n_l^2 + m_l^2 + z_l^2}\left(2 - 2\cos\left(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l\right)\right). \end{aligned} \tag{12}$$

Using the approximation $2 - 2\cos(x) \approx x^2$ [4] (12) simplifies

$$|C_l(\omega_1,\omega_2,\omega_3)|^2 \approx \frac{1}{n_l^2 + m_l^2 + z_l^2}\left(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l\right)^2. \tag{13}$$

We convert (13) to spherical frequency coordinates. The relationship between frequency and sampling yields $\omega_1 = 2\pi\Delta_x\rho\cos(\Phi)\cos(\Theta)$, $\omega_2 = 2\pi\Delta_y\rho\sin(\Phi)\cos(\Theta)$, and $\omega_3 = 2\pi\Delta_z\rho\sin(\Theta)$. Substituting these into (13) yields the following expression for $|C_l(\omega_1,\omega_2,\omega_3)|^2$

$$\begin{aligned} &\approx \frac{1}{n_l^2 + m_l^2 + z_l^2}(2\pi\rho)^2(n_l\Delta_x\cos(\Phi)\cos(\Theta) \\ &\qquad + m_l\Delta_y\sin(\Phi)\cos(\Theta) + z_l\Delta_z\sin(\Theta))^2 \end{aligned} \tag{14}$$

The local frequency response of the regularizer (2) is now

$$R_j(\rho,\Phi,\Theta) = (2\pi\rho)^2\kappa_j^2\sum_{l=1}^{N_l} r_j^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2, \tag{15}$$

where $e(\Phi,\Theta) \triangleq (\cos(\Phi)\cos(\Theta), \sin(\Phi)\cos(\Theta), \sin(\Theta))$, $\Delta \triangleq (\Delta_x, \Delta_y, \Delta_z)$, $\otimes$ is element-wise multiplication, and we assumed that $\kappa_j \approx \kappa_l$ for l within the neighborhood of j.

For the target response, R_o becomes

$$R_o(\rho,\Phi,\Theta) = (2\pi\rho)^2\kappa_o^2\sum_{l=1}^{N_l} r_o^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2, \tag{16}$$

where κ_o is the user-defined weights for target spatial resolution at the isocenter, and $\{r_o^l\}$ is the pre-defined directional weights, which determines the shape of the target response.

D. Regularization Design

Substituting (15) and (16) into (10) and simplifying yields

$$Q_j(\Phi,\Theta) \approx \frac{\kappa_0^2\tilde{w}_j(\Phi)}{\kappa_j^2\tilde{u}_o(\Phi)}Q_o(\Phi,\Theta), \tag{17}$$

where

$$Q_j(\Phi,\Theta) \triangleq \sum_{l=1}^{N_l} r_j^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2. \tag{18}$$

We solve the following weighted minimization problem to design the directional weighting coefficient vector $\boldsymbol{r}_j = (r_j^1, \cdots, r_j^{N_l})$ at the jth voxel

$$\begin{aligned} \boldsymbol{r}^j \triangleq \arg\min_{r^j\geq 0} &\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} D_w(\Phi,\Theta)|\breve{w}_j(\Phi,\Theta) \\ &- \sum_{l=1}^{N_l} r_l^j(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2|^2 d\Theta d\Phi, \end{aligned} \tag{19}$$

where the nonnegativity constraint ensures the regularizer's convexity and we define the modified weighting function

$$\breve{w}_j(\Phi,\Theta) \triangleq \frac{\kappa_0^2\tilde{w}_j(\Phi)}{\kappa_j^2\tilde{u}_o(\Phi)}\sum_{l=1}^{N_l} r_o^l(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2. \tag{20}$$

We choose $D_w = \cos(\Theta)$ to have more uniform distribution of sampled points. We view (19) as a weighted projection of

$\check{w}_j(\Phi)$ onto the space spanned by $\{[e(\Phi_l,\Theta_l)\otimes\Delta]^2\}$. Inserting the weight $\cos(\Theta)$ into the data-fitting part and expanding this term into 6 orthonormal basis functions, we can decompose $\sum_{l=1}^{N_l} r_j^l \cos^2(\Theta)(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2$ as $\boldsymbol{PTr}^j$, where $\boldsymbol{P}$ is an operator whose columns are the six orthonormal vectors, and $\boldsymbol{T}$ is a $6\times N_l$ linear combination matrix whose mth row is the following inner product

$$\boldsymbol{T}_{ml} = \frac{1}{2\pi^2}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2 p_m d\Theta d\Phi. \tag{21}$$

The orthonormal basis functions are given as follows

$$\begin{aligned}
p_1(\Phi,\Theta) &= \sqrt{\frac{8}{3}}\cos^2(\Theta)\\
p_2(\Phi,\Theta) &= \frac{16}{\sqrt{5}}\sin(\Phi)\sin(\Theta)\cos^3(\Theta)\\
p_3(\Phi,\Theta) &= \frac{16}{\sqrt{5}}\cos(\Phi)\sin(\Theta)\cos^3(\Theta)\\
p_4(\Phi,\Theta) &= \sqrt{\frac{96}{5}}\cos^2(\Theta)(\cos(2\Theta)-\frac{2}{3})\\
p_5(\Phi,\Theta) &= \frac{8}{\sqrt{35}}\cos(2\Phi)(1+\cos(2\Theta))\cos^2(\Theta)\\
p_6(\Phi,\Theta) &= \frac{32}{\sqrt{35}}\cos^4(\Theta)\cos(\Phi)\sin(\Phi),
\end{aligned}$$

and assuming $\Delta_x=\Delta_y$, the lth column of $\boldsymbol{T}$ is given by

$$\begin{bmatrix}
\sqrt{\frac{3}{2}}(\frac{1}{4}\Delta_x^2\cos^2(\Theta_l)+\frac{1}{3}\Delta_z^2\sin^2(\Theta_l))\\
\frac{\sqrt{5}}{5}\Delta_x\Delta_z\sin(\Phi_l)\sin(\Theta_l)\cos(\Theta_l)\\
\frac{\sqrt{5}}{5}\Delta_x\Delta_z\cos(\Phi_l)\sin(\Theta_l)\cos(\Theta_l)\\
-\frac{1}{\sqrt{30}}\Delta_z^2\sin^2(\Theta_l)\\
\frac{\sqrt{35}}{28}\Delta_x^2\cos^2(\Theta_l)(\cos^2(\Phi_l)-\sin^2(\Phi_l))\\
\frac{\sqrt{35}}{14}\Delta_x^2\cos(\Phi_l)\sin(\Phi_l)\cos^2(\Theta_l)
\end{bmatrix}.$$

Since $\boldsymbol{P}$ has orthonormal columns, we can represent the minimization problem (19) as the following simplified expression

$$\boldsymbol{r}^j = \underset{\boldsymbol{r}\geq 0}{\arg\min}\left\|\boldsymbol{Tr}-\boldsymbol{b}^j\right\|^2, \tag{22}$$

where $\boldsymbol{P}^*$ denotes the adjoint of $\boldsymbol{P}$ and $\boldsymbol{b}^j \triangleq \boldsymbol{P}^*\mathring{w}_j(\cdot)$, i.e., $b_k^j = 1/(2\pi^2)\int\int p_k(\Phi,\Theta)\mathring{w}_j(\Phi,\Theta)d\Phi d\Theta$ for $k=1,\cdots,6$, where $\mathring{w}_j(\Phi,\Theta) = \check{w}_j(\Phi,\Theta)\cos^2(\Theta)$. The minimization problem (22) is much smaller than (19). We solve (22) using NNLS algorithm [8].

The minimization problem (22) is under-determined and may have many different solutions $\boldsymbol{r}^j$ that are all global minima. Too many zeros in $\boldsymbol{r}^j$ may degrade the image since there will zeros in the Hessian [4]. To ensure that certain $\boldsymbol{r}^j$ values are greater than some small positive number $\boldsymbol{\epsilon}_l^j$, we modity (22) as follows

$$\begin{aligned}
\boldsymbol{r}^j &= \underset{\boldsymbol{r}\geq 0}{\arg\min}\left\|\boldsymbol{Tr}-(\boldsymbol{b}^j-\boldsymbol{T\epsilon}^j)\right\|^2 &(23)\\
&= \underset{\boldsymbol{r}\geq 0}{\arg\min}\left\|\boldsymbol{Tr}-\bar{\boldsymbol{b}}^j\right\|^2, &(24)
\end{aligned}$$

where $\bar{\boldsymbol{b}}^j \triangleq \boldsymbol{b}^j-\boldsymbol{T\epsilon}^j$. After minimization, we use the coefficients $\bar{\boldsymbol{r}}^j = \boldsymbol{r}^j+\boldsymbol{\epsilon}^j$ for our new regularizer (See [4] for a possible way to select $\boldsymbol{\epsilon}^j$).

III. Results

We simulated a 3rd-generation axial cone-beam CT system using the separable footprint projector [9]. The simulated system has $N_s = 888$ channels and $N_t = 64$ detector rows spaced by $\Delta_s = 1.0239$ mm and $\Delta_t = 1.09878$ mm, and 984 evenly spaced view angles over a 360 degree scan. The source to detector distance was 949 mm, and the source to rotation center distance was 541 mm. We included a quarter detector offset in the channel direction to reduce aliasing. The XCAT phantom [10] was used, and the image was reconstructed to a $512\times512\times122$ grid with pixel size $\Delta_x=\Delta_y=0.9766$ mm and $\Delta_z = 0.625$ mm. Poisson noise was added to the sinogram, and the statistical weighting was chosen as $w_i = \exp(-[\boldsymbol{Ax}]_i)$. The regularization parameter β was selected such that the target PSF at the isocenter has a full-width at half-maximum (FWHM) of approximately 1.4 mm in xy and 0.9 mm in z.

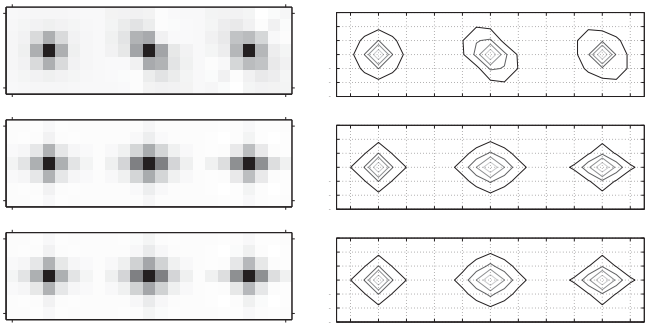


Fig. 1. Impulse responses of conventional regularization (middle column) and proposed regularization (right column) at (-66,217,-17) (mm), which is a fully sampled location. Target impulse response is given as a reference (left column). Each row corresponds to xy, xz, and yz profiles, respectively. Each contour was plotted based on its own peak value.

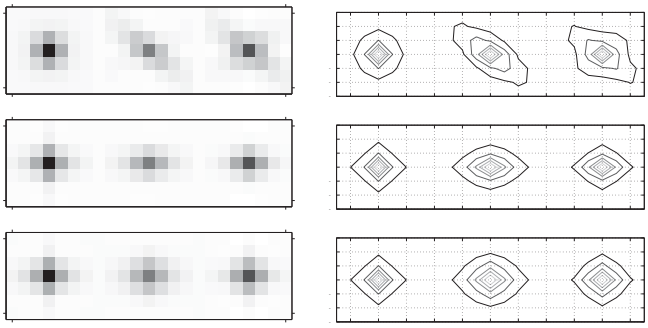


Fig. 2. Impulse responses of conventional regularization (middle column) and proposed regularization (right column) at (-117,-67,17) (mm), which is a insufficiently sampled location. Target impulse response is given as a reference (left column). Each row corresponds to xy, xz, and yz profiles, respectively. Each contour was plotted based on its own peak value.

Figs. 1 and 2 compare impulse responses of conventional regularization and proposed regularization for two different voxels with different sampling properties. There is a considerable anisotropy at both locations, especially for voxels with insufficient sampling. The main reasons for the anisotropy are statistical weighting and scan geometry. The spatial resolution of the voxel in Fig. 1 is primarily affected by the statistical weights, and our proposed method gives more isotropic

impulse response. The location in Fig. 2 is greatly affected by scan geometry, and our proposed method achieves limited improvements.

Fig. 3 compares reconstructed images with various methods. Iteratively reconstructed images show better noise characteristics compared to the FDK reconstruction, but they may have more anisotropic spatial resolution especially at the voxels with less samplings. The true image blurred by the target impulse response was provided as a reference to assess the improvements of our proposed method. In Fig. 4 closely compares the reconstructed images with conventional regularization and the proposed regularization. Overall, the reconstructed image with the proposed regularizer has better resolution characteristics, but has slightly more noise. At locations indicated by the arrows, the proposed regularization shows noticeable improvements (better match to target).

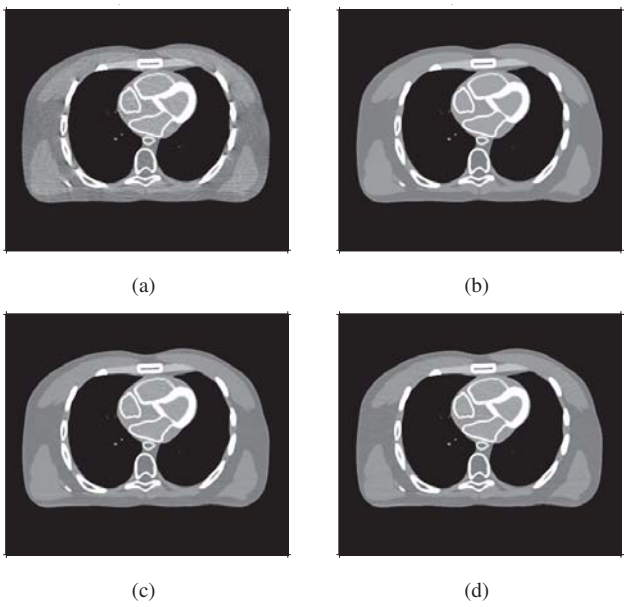


Fig. 3. Reconstructed images at end slice (a) FDK reconstruction (b) True image blurred by the target impulse response (c) Iterative reconstruction with conventional regularizer (d) Iterative reconstruction with designed regularizer

IV. Discussion

We proposed a regularization design method for 3D axial CT that aims to improve resolution uniformity and isotropy. The proposed regularization showed improved spatial resolution characteristics compared to the conventional regularization for the full scan geometry. However, the designed impulse responses do not match the target response precisely and locations with insufficient sampling still suffer from anisotropic resolution. Since 3D axial short scans can suffer from severe anisotropy at the end slices due to their scan geometry, the proposed method may have difficulties achieving desired isotropic resolutions for short scans. We hope to compensate for this with improved regularization design. Furthermore, since edge-preserving regularization is mostly used in practice instead of the quadratic regularization, we will investigate using the designed directional weights in edge-preserving regularization. Our future work will address these issues and focus on refining the method to obtain better spatial uniformity for different scan geometries and to further improve the computational efficiency of the method.

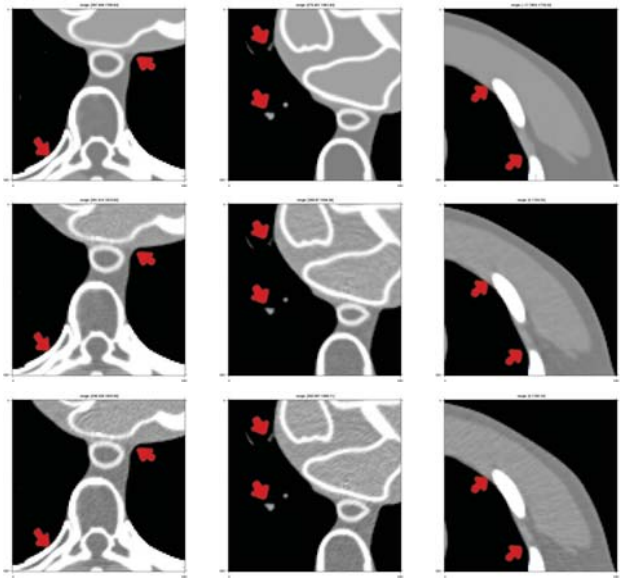


Fig. 4. Reconstructed images with conventional regularization (middle row) and proposed regularization (bottom row) at different locations on end slices. True image blurred by the target impulse response is given as a reference (top row).

Acknowledgement

The authors would like to thank J. B. Thibault and Debashish Pal for their valuable comments.

References

[1] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[2] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Im. Proc.*, vol. 5, no. 9, pp. 1346–58, Sep. 1996.

[3] J. W. Stayman and J. A. Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 269–84, Mar. 2004.

[4] H. R. Shi and J. A. Fessler, "Quadratic regularization design for 2D CT," *IEEE Trans. Med. Imag.*, vol. 28, no. 5, pp. 645–56, May 2009.

[5] H. Shi and J. A. Fessler, "Quadratic regularization design for 3D axial CT," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2006, pp. 2834–6.

[6] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, May 2000.

[7] S. Schmitt and J. A. Fessler, "Fast variance computation for quadratically penalized iterative reconstruction of 3D axial CT images," in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2012, pp. 3287–92.

[8] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Prentice-Hall, 1974.

[9] Y. Long, J. A. Fessler, and J. M. Balter, "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1839–50, Nov. 2010.

[10] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–8, Aug. 2008.

CT Reconstruction of 3D Wavelet Coefficients and its Application to Nondestructive Testing

Steven Oeckl*

***Abstract*—This paper presents a new approach for multiresolution reconstruction in cone-beam tomography. The approximative inverse for feature reconstruction is used to derive filter kernels for reconstructing the 3D wavelet approximation and wavelet detail coefficients directly from the projection data. Beyond the reconstruction kernels, applications of multiresolution reconstruction in the field of nondestructive testing are shown: The algorithm supports progressive reconstruction and local tomography for recovering only a region of interest inside the investigated volume. The features of the reconstruction algorithm are shown by means of simulation data as well as real data of an aluminium casting from automobile industry**

***Index Terms*—CT, reconstruction, approximate inverse, wavelet, multiresolution, nondestructive testing, merging of image processing and reconstruction, data compression**

I. Introduction

Since wavelet analysis has become a powerful tool for signal and image processing, the multiresolution approach provides a solution for many practical applications. In X-ray computerized tomography (CT) algorithms for multiresolution two dimensional (2D) parallel beam, 2D fan-beam, and three dimensional cone-beam (Feldkamp-type) reconstruction using tensor or quincunx wavelets were introduced in [1]–[3]. These reconstruction formulas are based on the strong relationship between the continuous wavelet transform and the Radon transform as mentioned in [4].

In this contribution we use a different approach to achieve an algorithm for reconstructing an object at different resolutions: The approximate inverse, introduced in [5], is a method for solving first kind operator equations $Af = g$ in a stable way. Instead of determining the exact solution f, an inversion operator for f_e is calculated, where f_e is associated to f via the inner product $\langle f, e\rangle$ using a mollifier e. Applying the approximate inverse to CT yields a reconstruction algorithm of filtered backprojection type [6]. In [7] the concept of approximate inverse was extended for calculating a feature of f represented as an appropriate operator T directly from the measured data g. In this case the inner product $\langle Tf, e\rangle$ is calculated. Choosing the operator T as the discrete wavelet transform yields a filtered backprojection algorithm for reconstructing the wavelet coefficients of f.

We start with a short introduction to the concept of approximate inverse in section II. Necessary basics on nonseparable multiresolution analysis are given in section III. Section IV provides the definition of the X-ray transform and some relevant results. The determination of filter kernels for calculating the 3D wavelet and approximation coefficients directly from the measured projection data is presented in section V. We show the results of our reconstruction method in section VI and conlude this contribution with a summary and an outlook in section VII.

* Process Integrated Inspection Systems, A Dept. of the Fraunhofer Institute for Integrated Circuits, Fürth, Germany
Corresponding author: Steven.Oeckl@iis.fraunhofer.de

II. Approximative Inverse

For separable hilbert spaces U and V we denote the space of linear continuous operators from U to V by $L(U,V)$. The range of $\mathcal{A} \in L(U,V)$ is defined by $R(\mathcal{A})$ and we write $\mathcal{A}^*$ for the adjoint of $\mathcal{A}$.

Let $\mathcal{A} \in L(U,V)$ and $e \in R(\mathcal{A}^*)$. An element $\kappa_{\mathcal{A}}(e) \in V$ with

$$\mathcal{A}^* = e \tag{1}$$

is called *reconstruction kernel for $\mathcal{A}$ (concerning e)*.

Using the reconstruction kernel we can define the approximate inverse as follows: Let $\Omega, \Omega' \subset \mathbb{R}^n$ be bounded domains, $U := L^2(\Omega')$, V a separable hilbert space, $\mathcal{A} \in L(U,V)$, $g \in V$ and $e \in L^2(\Omega \times \Omega')$ fulfilling $e(x,\cdot) \in R(\mathcal{A}^*)$ for all $x \in \Omega$. The operator

$$\mathcal{S}_{\mathcal{A}}(e)g(x) := \langle g, \kappa_{\mathcal{A}}(e(x,\cdot))\rangle_V\,, \quad x \in \Omega\,, \tag{2}$$

is called *approximate inverse of $\mathcal{A}$ (concerning e)*.

Assume the first kind operator equation $\mathcal{A}f = g$. The idea of the approximate inverse is to calculate an approximation $\langle f, e(x,\cdot)\rangle_U$ of the solution instead of the exact solution f. This idea can be seen using the above definitions

$$\begin{aligned}\langle f, e(x,\cdot)\rangle_U &= \langle f, \mathcal{A}^*\kappa_{\mathcal{A}}(e(x,\cdot))\rangle_U \\ &= \langle \mathcal{A}f, \kappa_{\mathcal{A}}(e(x,\cdot))\rangle_V \\ &= \langle g, \kappa_{\mathcal{A}}(e(x,\cdot))\rangle_V \\ &= \mathcal{S}_{\mathcal{A}}(e)g(x)\,. \end{aligned} \tag{3}$$

Since we are interested in a certain features of f, i.e. the wavelet coefficients, instead of the solution of $\mathcal{A}f = g$ itself, we extend the above concept to this situation, see [7]. Let $\Omega, \Omega', \Omega'' \subset \mathbb{R}^n$ be bounded domains, $U := L^2(\Omega')$, V a separable hilbert space, $\mathcal{A} \in L(U,V)$, $W := L^2(\Omega'')$, $\mathcal{T} \in L(U,W)$, $g \in V$ and $e \in L^2(\Omega \times \Omega'')$ fulfilling $\mathcal{T}^*e(x,\cdot) \in R(\mathcal{A}^*)$ for all $x \in \Omega$. The operator

$$\mathcal{S}_{\mathcal{A}}(e,\mathcal{T})g(x) := \langle g, \kappa_{\mathcal{A}}(\mathcal{T}^*e(x,\cdot))\rangle_V\,, \quad x \in \Omega,$$

is called *approximative inverse of $\mathcal{A}$ (concerning e) for (calculating the property) $\mathcal{T}$*. This definition is again motivated by

the following relation

$$\begin{aligned}\langle \mathcal{T}f, e(x,\cdot)\rangle_W &= \langle f, \mathcal{T}^* e(x,\cdot)\rangle_U \\ &= \langle f, \mathcal{A}^*\kappa_{\mathcal{A}}(\mathcal{T}^* e(x,\cdot))\rangle_U \\ &= \langle \mathcal{A}f, \kappa_{\mathcal{A}}(\mathcal{T}^* e(x,\cdot))\rangle_V \\ &= \langle g, \kappa_{\mathcal{A}}(\mathcal{T}^* e(x,\cdot))\rangle_V \\ &= \mathcal{S}_{\mathcal{A}}(e,\mathcal{T})g(x)\,.\end{aligned}$$

Instead of calculating the exact feature $\mathcal{T}f$ we use the extended approximate inverse to determine an approximation $\langle \mathcal{T}f, e(x,\cdot)\rangle_W$ using again an appropriate mollifier e.

In the context of CT the calculation of reconstruction kernels using the classical approximate inverse is well known, see [6]. The following result shows the relation between the classical and the extended approximate inverse for properties. This relation will lead us to efficient algorithms for calculating reconstruction kernels for the extended approximate inverse. Let U, V, X and Y be separable hilbert spaces, $\mathcal{A} \in L(U,V)$ and $\mathcal{B} \in L(X,Y)$. In addition let $\mathcal{T}^1 \in L(X,U)$ and $\mathcal{T}^2 \in L(Y,V)$ fulfilling

$$\mathcal{A}\mathcal{T}^1 = \mathcal{T}^2\mathcal{B} \tag{4}$$

and let be $e \in R(\mathcal{A}^*)$. Then ${\mathcal{T}^1}^* e \in R(\mathcal{B}^*)$ and

$$\kappa_{\mathcal{B}}({\mathcal{T}^1}^* e) = {\mathcal{T}^2}^* \kappa_{\mathcal{A}}(e)\,. \tag{5}$$

To proof this we can write

$${\mathcal{T}^1}^* e = {\mathcal{T}^1}^* \mathcal{A}^* \kappa_{\mathcal{A}}(e) = \mathcal{B}^* {\mathcal{T}^2}^* \kappa_{\mathcal{A}}(e)\,.$$

III. Multiresolution Analysis

For calculating the wavelet coefficients of f directly from the projection data using the approximate inverse we introduce first some results from multiresolution analysis, for details see [8], [9]. Let $D \in \mathbb{Z}^{nxn}$ be a dilation matrix and define $M := |detD|$. Let $\{\phi, \tilde{\phi}\}$ be a pair of dual scaling functions of a n-dimensional multiresolution analysis, and let $\{\psi^i, \tilde{\psi}^i\}$, $i = 1,\ldots,M$, be $M-1$ pairs of dual mother wavelets. For a fixed $J \in \mathbb{Z}$ and $f \in L^2(\mathbb{R}^n)$ we have the wavelet expansion

$$f = \sum_{k\in\mathbb{Z}^n} \langle f, \phi_{J,k}\rangle \tilde{\phi}_{J,k} + \sum_{i=1}^{M-1}\sum_{j<J}\sum_{k\in\mathbb{Z}^n} \langle f, \psi^i_{j,k}\rangle \tilde{\psi}^i_{j,k}\,,$$

where $\phi_{j,k} := |detD|^{-(j/2)}\,\phi(D^{-j}\cdot - k)$. We call $\langle f, \psi^i_{j,k}\rangle$ and $\langle f, \phi_{J,k}\rangle$ wavelet and approximation coefficients respectively.

In [1] it was shown that the wavelet and approximation coefficients are the result of appropriate convolution operators and therefore can be calculated using the approximate inverse for properties. Let $\Omega^n_r := \{x \in \mathbb{R}^n : \|x\| < r\}$ be the open ball in $\mathbb{R}^n$ with radius $r \in \mathbb{R}$. For $r_1, r_2 \in \mathbb{R}^+$, $f \in L^2(\Omega^n_{r_1})$ and $\psi \in L^2(\mathbb{R}^n)$ we have $(f * \psi) := \int_{\mathbb{R}^n} f(x)\psi(\cdot - x)\,dx \in L^2(\Omega^n_{r_2})$. The operator

$$\begin{aligned}\mathcal{C}_\psi : L^2(\Omega^n_{r_1}) &\to L^2(\Omega^n_{r_2}) \\ f &\mapsto \mathcal{C}_\psi f := (f * \psi)\end{aligned}$$

is called *convolution with* ψ. Let $\breve{\psi} := \overline{\psi(-\cdot)}$, $j \in \mathbb{Z}$ and $D \in M_n(\mathbb{Z})$ a dilatation matrix. Then for all $k \in \mathbb{Z}^n$ with $(D^j)^{-1}k \in \Omega^n_{r_2}$ holds

$$\begin{aligned}\langle f, \psi_{j,k}\rangle_{L^2(\mathbb{R}^n)} &= \int_{\mathbb{R}^n} f(t)\overline{\psi(D^j t - k)}\,dt \\ &= \int_{\mathbb{R}^n} f(t)\breve{\psi}(k - D^j t)\,dt \\ &= \int_{\mathbb{R}^n} f(t)\breve{\psi}(D^j((D^j)^{-1}k - t))\,dt \\ &= \int_{\mathbb{R}^n} f(t)\breve{\psi}_{j,0}((D^j)^{-1}k - t)\,dt \\ &= \mathcal{C}_{\breve{\psi}_{j,0}} f((D^j)^{-1}k)\,.\end{aligned}$$

The corresponding equation holds for the approximation coefficients. Therefore we can determine the coefficients for the wavelet expansion directly from the measured data g of the operator equation $\mathcal{A}f = g$ by using the approximate inverse of $\mathcal{A}$ for $\mathcal{C}_{\breve{\psi}_{j,0}}$.

IV. X-ray transform

The X-ray transform is the mathematical model of cone beam CT. In this section we restrict our investigations to the three dimensional (3D) case $n = 3$. We denote the unit sphere in $\mathbb{R}^3$ with S^3. For $f \in L^1(\mathbb{R}^3)$, $a \in \mathbb{R}^3$, $\theta \in S^2$ we define

$$\mathcal{D}f(a,\theta) := \mathcal{D}_a f := \int_0^\infty f(a + \rho\theta)\,d\rho\,.$$

The operator $\mathcal{D}$ is called *X-ray transform*. Next we consider the X-ray transform as a continuous linear operator concerning an appropriate curve. Therefore let $\Lambda \subset \mathbb{R}$ be a closed interval, $\phi : \Lambda \to \mathbb{R}^3$ a curve and $\Gamma_\phi := R(\phi) \subset \mathbb{R}^3$ the range of ϕ. The set of all curves with a range outside of Ω^3_r is denoted by $\Phi^{3,r}(\Lambda) := \{\phi : \Lambda \to \mathbb{R}^3 \,|\, R(\phi) \subset \mathbb{R}^3 \setminus \Omega^3_r\}$. We call $\phi \in \Phi^{n,r}(\Lambda)$ a *Tuy curve*, if ϕ is bounded, continuous, differentiable almost everywhere and if for all $(x,\theta) \in \Omega^n_r \times S^{n-1}$ there exists a $\lambda \in \Lambda$ such that

$$\langle x, \theta\rangle = \langle \phi(\lambda), \theta\rangle \tag{6}$$

and

$$\langle \phi'(\lambda), \theta\rangle \neq 0\,. \tag{7}$$

Let $\phi \in \Phi^{n,r}(\Lambda)$ be a Tuy curve. Then the X-ray transform $\mathcal{D}$ is a continuous linear mapping

$$\mathcal{D} : L^2(\Omega^n_r) \to L^2(\Gamma_\phi \times S^3)\,.$$

The adjoint operator $\mathcal{D}^* : L^2(\Gamma_\phi \times S^3) \to L^2(\Omega^3_r)$ is given as

$$\mathcal{D}^* g(x) := \int_\Lambda \|x - \phi(\lambda)\|^{-2} g\left(\frac{x - \phi(\lambda)}{\|x - \phi(\lambda)\|}\right) d\lambda\,.$$

The inversion formula in [10] is based on the adjoint X-ray transform and is therefore suitable for calculating reconstruction kernels. The formula reads

$$f = \frac{1}{8\pi^2}\mathcal{D}^*\mathcal{T}\mathcal{M}_{\Gamma_\phi}\mathcal{T}\mathcal{D}f\,, \tag{8}$$

where for $\lambda \in \Lambda$, $\omega \in S^2$ and $s \in \mathbb{R}$

$$\mathcal{T}g(\phi(\lambda),\omega) := \int_{S^2} g(\phi(\lambda),\theta)\delta'(\langle\theta,\omega\rangle)\,,$$

$$\mathcal{M}_{\Gamma_\phi} h(\phi(\lambda),\omega) := |\langle\phi'(\lambda),\omega\rangle| m_\phi(\omega,\langle\phi(\lambda),\omega\rangle) h(\phi(\lambda),\omega)$$

and $m_\phi(\omega, \langle\phi(\lambda), \omega\rangle) := n_\phi(\omega, \langle\phi(\lambda), \omega\rangle)^{-1}$ the inverse of the *Crofton symbol*

$$n_\phi(\omega, s) := \#\{\lambda \in \Lambda : \langle\phi(\lambda), \omega\rangle = s\}\,.$$

For an efficient calculation of reconstruction kernels for computing the wavelet coefficients directly from the measured projection data the following relation between the X-ray transform and the convolution operation is important. We have

$$\begin{aligned}
\mathcal{D}_{\phi(\lambda)}\mathcal{C}_g f(\theta) &= \int_0^\infty \mathcal{C}_g f(\phi(\lambda) + t\theta)\, dt \\
&= \int_0^\infty \mathcal{C}_f g(\phi(\lambda) + t\theta)\, dt \\
&= \int_0^\infty \int_{\Omega_r^n} g(x) f(\phi(\lambda) + t\theta - x)\, dx\, dt \\
&= \int_{\Omega_r^n} g(x) \int_0^\infty f(\phi(\lambda) - x + t\theta)\, dt\, dx \\
&= \int_{\Omega_r^n} g(x) \mathcal{D}_{\phi(\lambda)-x} f(\theta)\, dx \\
&= \mathcal{C}_g(\mathcal{D}_{(\cdot)} g(\theta))(\phi(\lambda))\,. \qquad (9)
\end{aligned}$$

V. Reconstruction kernels

The approximate inverse for determine a smoothed solution of $\mathcal{D}f = g$ reads

$$\begin{aligned}
\langle f, e(x,\cdot)\rangle_U &= \mathcal{S}_\mathcal{D}(e)\mathcal{D}f(x) \\
&= \int_\Lambda \int_{S^2} \mathcal{D}_{\phi(\lambda)} f(\theta) \kappa_\mathcal{D}(e(x,\cdot))(\phi(\lambda), \theta)\, d\theta\, d\lambda\,, \qquad (10)
\end{aligned}$$

where

$$\kappa_\mathcal{D}(e(x,\cdot)) = \mathcal{T}\mathcal{M}_{\Gamma_\phi}\mathcal{T}\mathcal{D}e(x,\cdot)\,.$$

This follows immediately from the inversion formula (8)

$$\mathcal{D}^*\kappa_\mathcal{D}(e(x,\cdot)) = \mathcal{D}^*\mathcal{T}\mathcal{M}_{\Gamma_\phi}\mathcal{T}\mathcal{D}e(x,\cdot) = e(x,\cdot)\,.$$

Due to relation (9) the assumption (4) is fulfilled. Therefore we can apply (5) to determine the reconstruction kernel for calculating the wavelet coefficients $\mathcal{C}^*_{\check{\psi}_{j,0}} f$ directly from the measured data as follows

$$\kappa_\mathcal{D}(\mathcal{C}^*_{\check{\psi}_{j,0}} e(x,\cdot))(\phi(\lambda), \theta) = \mathcal{C}^*_{\check{\psi}_{j,0}}(\kappa_\mathcal{D}(e(x,\cdot))(\cdot,\theta))(\phi(\lambda))\,.$$

The extended approximate inverse for determine the wavelet coefficients $\mathcal{C}^*_{\check{\psi}_{j,0}} f$ reads

$$\begin{aligned}
&\langle \mathcal{T}f, e(x,\cdot)\rangle_W = \mathcal{S}_\mathcal{A}(e, \mathcal{C}_{\check{\psi}_{j,0}})\mathcal{D}f(x) \\
&= \int_\Lambda \int_{S^2} \mathcal{D}_{\phi(\lambda)} f(\theta) \kappa_\mathcal{D}(\mathcal{C}^*_{\check{\psi}_{j,0}} e(x,\cdot))(\phi(\lambda), \theta)\, d\theta\, d\lambda\,. \qquad (11)
\end{aligned}$$

In [6] it was shown that in case of a circle trajectory formula (10) provides a stable reconstruction algorithm, although a circle is not a Tuy curve. We use also a circle trajectory because this curve is the most common curve in industrial CT. If the mollifier e in this case is translation invariant, then the inner integral concerning S^2 in (10) acts approximately as a convolution, see [6]. The same holds for the inner integral concerning S^2 in (11). Therefore we get an algorithm of filtered backprojection type for reconstructing the wavelet and approximation coefficients $\mathcal{C}^*_{\check{\psi}_{j,0}} f$.

VI. Results

This section is divided into two parts. First we present some results using our algorithm on synthetic data. Second we would like to show that the approach of calculating a 3D multiresolution analyis directly from the projection data provides some benefits in practical applications.

In Figure 1 we show a single slice of a 3D wavelet and approximation coefficient reconstruction using a coiflet and the 3D extension of the quincunx dilatation matrix. The full cone angle was two times $7,78$ degrees and 400 projections consisting of 256x256 pixels were used to reconstruct a volume data set with 256x256x256 voxels. We choose the gaussian with standard deviation $\sigma \approx \sqrt{2}$ as mollifier.

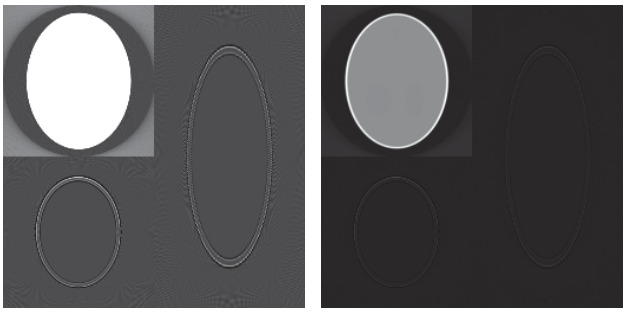


Fig. 1. Reconstruction of 3D wavelet and approximation coefficients using a coiflet and the 3D extension of the quincunx dilatation matrix. The full cone angle was two times $7,78$ degrees and 400 projections consisting of 256x256 pixels were used to reconstruct a volume data set with 256x256x256 voxels. We choose the gaussian with standard deviation $\sigma \approx \sqrt{2}$ as mollifier. Left: Grey value range of $[-0, 1, 0.25]$. Right: Grey value range of $[-0.1, 1.0]$.

Using the same setup as in the situation of figure 1 we show in figure 2 a comparison between calculating the approximation coefficients (without downsampling) in two steps and a reconstruction of the approximation coefficients directly from the projection data. A two step calculation means performing first a standard reconstruction followed be a 3D low pass filtering of the volume data set to achieve the approximation coefficients related to the coiflet.

Fig. 2. Comparison between calculating the approximation coefficients (without downsampling) in two steps and a reconstruction of the approximation coefficients directly from the projection data using the same setup as in figure 1. Left: Two step calculation, i.e. performing first a standard reconstruction followed be a 3D low pass filtering of the volume data set; grey value range $[0.49, 0, 535]$. Center: Reconstruction of the approximation coefficients directly from the projection data using the extended approximate inverse; grey value range $[0.49, 0, 535]$. Right: Absolute difference between left and right image; grey value range $[0.0, 0.005]$.

The possibility to perform a progressive reconstruction is the main advantage of multiresolution tomographic reconstruction for practical applications. Reconstructing the approximation

coefficients at a high decomposition level yields a first impression of the specimen. After selecting a region of interest within the approximation only the detail coefficients of the selected region plus a certain border must be reconstructed to achieve high resolution inside the region of interest. In non-destructive testing progressive reconstruction can be used to incorporate previous knowledge about the specimen, the scanning geometry, and the inspection task efficiently into the reconstruction method.

In figure 3 we show the result of merging CT reconstruction and image processing for automatic defect detection in aluminium castings. After reconstructing the approximation coefficients the potential defect areas can be determined automatically. Afterwards only these areas need to be reconstructed at highest resolution the perform an exact defect characterisation. This approach reduces the amount of data to be processed and therefore speeds up the casting inspection.

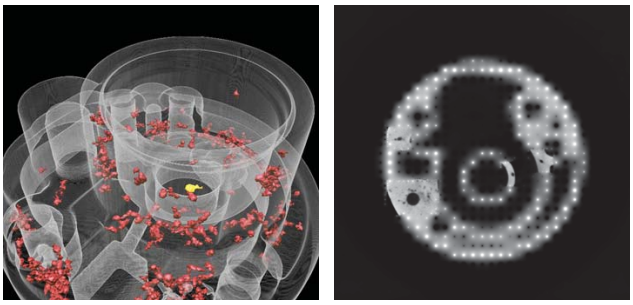


Fig. 3. Merging of reconstruction and image processing speeds up the process for casting inspection. Only the defect areas need to be reconstructed at highest resolution. Left: 3D representation of casting inspection result. Right: 3D multiresolution reconstruction where only the relevant regions are reconstructed at highest resolution.

A 3D multiresolution reconstruction enables data compression during the reconstruction process. For example a wavelet thresholding can be performed after every reconstructed decomposition level. This speeds up the data transfer and therefore the reconstruction time in case of huge volume data sets processed using cluster computing. An example can be seen in figure 4.

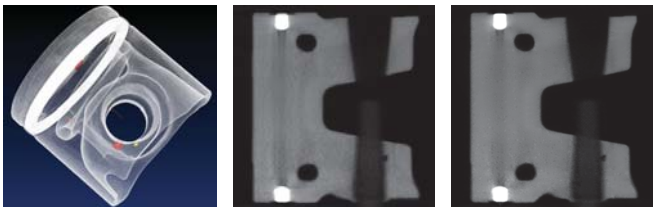


Fig. 4. Data compression during the reconstruction step. Left: 3D representation of a huge data set with casting defects inside. Center: Standard reconstruction using an industrial CT setup. Right: Result of compressed CT reconstruction using 3D multiresolution analysis with an compression rate of 13.34 %. All relevant structures for automatic evaluation tasks including the casting defect in the lower right quadrant of the image are still visible.

Also Region-Of-Interest (ROI) reconstruction using multi-scan approaches is an application field of 3D multiresolution reconstruction.

VII. Conclusion and Future Work

In this contribution we show an approach for reconstructing 3D wavelet and approximation coefficients directly from the projection data. We use the extended approximate inverse for calculating properties to derive reconstruction kernels for a filtered backprojection for a circle trajectory. Beyond results on synthetic data we present also practical applications in industrial CT.

Next steps will be the expansion of the kernel determination to other trajectories, for example the helical trajectory, and to other properties than wavelet coefficients. For metrology applications the reconstruction of object edges instead of absorption coefficients should provide some benefits.

Acknowledgment

The author would like to thank Tobias Schön for many fruitful discussions and support in implementation tasks. The author would like to thank also Roland Gruber, Balthasar Reuter and Florin Schwappach for their support in implementation tasks.

References

[1] S. Bonnet, F. Peyrin, F. Turjman, and R. Prost, "Multiresolution reconstruction in fan-beam tomography," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 169–176, March 2002.

[2] ——, "Nonseparable wavelet-based cone-beam reconstruction in 3-d rotational angiography," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 360–367, March 2003.

[3] F. Rashid-Farrokhi, K. J. R. Liu, C. A. Berenstein, and D. Walnut, "Wavelet-based multiresolution local tomography," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1412–1430, October 1997.

[4] D. Walnut, "Applications of gabor and wavelet expansions to the radon transform," in *Probabilistic and Stochastic Methods in Analysis, With Applications*, J. S. B. et al, Ed. Kluwer Academic Publishers, 1992, pp. 187–205.

[5] A. K. Louis, "Approximate inverse for linear and some nonlinear problems," *Inverse Problems*, vol. 12, pp. 175–190, 1996.

[6] ——, "Filter design in three-dimensional cone beam tomography: circular scanning geometry," *Inverse Problems*, vol. 19, pp. 31–40, 2003.

[7] A. Louis, "Combining image reconstruction and image analysis with an application to two-dimensional tomography," *SIAM Journal on Imaging Sciences*, vol. 1, no. 2, pp. 188–208, 2008.

[8] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.

[9] A. K. Louis, P. Maaß, and A. Rieder, *Wavelets*. Teubner, 1998.

[10] A. Louis, "Development of algorithms in computerized tomography," in *The Radon Transform, Inverse Problems, and Tomography*, 2006, pp. 25–42.

Constrained TV-minimization Reconstruction in Low-dose Breast CT

Junguo Bian, Kai Yang, Xiao Han, Emil Y. Sidky, John M. Boone, and Xiaochuan Pan

Abstract—**In this work, we perform a preliminary study on image reconstruction by using optimization-based algorithms from low-dose data of patients collected with a dedicated breast CT scanner. The reconstruction is formulated as a constrained minimization program, and the adaptive steepest descend and projection onto convex sets (ASD-POCS) is used to solve the program. Because the data were collected with low-dose exposure comparable to that in a typical two-view mammography examination, they have a relatively low signal-to-noise ratio (SNR), and reconstruction from them is challenging in terms of revealing detailed information of clinical value. Based upon the reconstruction, we also conduct characterization studies in terms of visualization and image-power spectra. The results of our study suggest that, constrained TV-minimization-based reconstruction may yield breast images with improvement over those obtained with the standard FDK algorithm.**

I. Introduction

Breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer mortality among women in the US. Death rates from breast cancer have been steadily decreasing since 1990, largely because of the early detection of breast cancer. Mammography has played a major role in early breast cancer detection. However, mammography has inherent limit because mammography provides only 2D projection images of a 3D breast, which yields overlapping structures, and thus can lead to false-negative and/or false-positive detections, especially for dense breasts. Researchers are thus actively developing 3D breast-imaging techniques, including breast tomosynthesis and dedicated breast CT, to overcome this inherent limitations of conventional 2D mammography. Tomosynthesis can be interpreted as a CT method acquiring data from a small number of views over a limited angular range. Although it can remove some overlapping artifacts in mammography, it yields non-isotropic-resolution breast images with poor out-of-plane (plane perpendicular to the detector plane) resolution. Breast CT was suggested in the 1970s [1]–[3] and has recently received renewed interest [4]–[7].

Current dedicated breast-CT prototypes are based on flat-panel detectors and generally adopt a circular cone-beam imaging configuration. Analytic-based algorithms such as the FDK algorithm [8] are currently used for image reconstruction for breast-CT prototypes. However, analytic-based algorithms require a large number of densely sampled views. For example, current breast-CT scans collect data typically at $\sim$500 views for avoiding sampling artifacts in images reconstructed by use of analytic-based algorithms. Because the total imaging dose delivered to the patient in a breast-CT scan remains about the same as that in a typical two-view mammography examination, the use of a large number of views can lead to data with low signal-to-noise ratio (SNR) and images with high noise. The soft tissues of a breast have similar attenuation properties to water and glandular tissues of a breast have very fine structures. Low contrast, fine structure, and low-SNR projection data thus makes reconstruction improvement from low-dose dedicated breast-CT data very challenging.

There exists increased interest in development and evaluation of optimization-based algorithms for image reconstruction in cone-beam CT (CBCT), because, when applied to the large amount of data typically collected in current applications, they may yield images with improved quality over that of reconstructions obtained by use of analytic-based algorithms such as FDK-based algorithms, and because they can be more flexible for accommodating imaging conditions of practical significance than analytic-based algorithms. A great deal of results have been reported on image reconstruction by use of optimization-based algorithms from data acquired with CBCT scanners [9]–[15]. Among optimization-based algorithms developed recently, algorithms that exploiting image-sparsity properties have attracted considerable attention as their potential to reconstruct images of practical utility has been demonstrated. The adaptive-steepest-descent-projection-onto-convex-set (ASD-POCS) algorithm is one of such algorithms, and it has been applied to reconstructing images from data collected with an array of scanning configurations. Evaluation studies [11], [12], [16], [17] have shown a potential of the ASD-POCS algorithm in yielding quality images from data less than what are needed by an analytic-based algorithm in current imaging applications.

In this work, we investigate optimization-based image reconstruction from low-dose breast CT patient data by using ASD-POCS algorithm. The focus of this work is to investigate, under challenging low-dose condition, whether ASD-POCS algorithm can yield any improvement over breast-CT images reconstructed currently with the FDK algorithm. The main purpose of this study is not to develop new reconstruction algorithms, but rather to characterize an existing optimization-based algorithm, and demonstrate the application of the algorithm under practical challenging conditions. Although it is important to develop new algorithms, it is actually more important to characterize and tailor "new" or "existing" algorithms to address problems of practical significance and identify candidates, parameters that may be useful for practical applications.

J. Bian, X Han and E. Y. Sidky are with Department of Radiology, the University of Chicago.

K. Yang and J. Boone are with Department of Radiology, University of California, Davis.

X. Pan is with Departments of Radiology and Radiation & Cellular Oncology, the University of Chicago.

II. Materials and Methods

A. The breast-CT prototype and data acquisition

The prototype breast-CT scanner used in this study consists of an X-ray source with a tungsten anode, a flat-panel detector, and an integrated bearing-motor-encoder system. The detector has 1024×768 elements, with a pitch of 0.0388 cm. The source-to-detector and source-to-iso-center distances are about 87.78 cm and 45.83 cm respectively. For patient studies, the subjects were set up in a prone position, and their breasts hung pendant through a hole in the table top. The source and detector rotate around the patient breasts.

The patient data were acquired as part of an on-going clinical trial at UC-Davis medical center for evaluating the clinical utility of the breast CT prototype. For each data set, cone-beam projections at 500 views uniformly distributed over 2π were acquired in about 17 seconds. The patients were advised to hold their breath during the scan. The projection data are corrected for certain physical factors, including detector-pixel defects, scatter and beam-hardening, before image reconstruction.

B. Optimization-based image reconstruction

As discussed previously [16], [17], there are three import components for optimization-based image reconstruction, which are imaging model, optimization program, and optimization-based algorithm for solving the optimization program, which we discuss briefly below.

1) Imaging model: We use a linear discrete system model for modeling the CBCT imaging process in optimization-based reconstruction:

$$\mathbf{g_0} = \mathcal{H}\mathbf{f}, \tag{1}$$

where $\mathcal{H}$ is an system matrix modeling the cone-beam X-ray transform; vectors $\mathbf{g_0}$ and $\mathbf{f}$ denote discrete model data and discrete image to be reconstructed. Each entry of $\mathbf{g_0}$ or $\mathbf{f}$ denotes a model-data value within an detector element or an image value within a voxel. In practice, we only have measured data $\mathbf{g}$ instead of model data $\mathbf{g_0}$, which necessarily contains components not consistent with the linear system. Reconstruction of image $\mathbf{f}$ is equivalent to inverting the linear system in Eq. (1) from knowledge of measured, discrete data.

2) Reconstruction program and algorithms: Based upon Eq. (1), an optimization program has been developed [9]–[12] for CBCT-image reconstruction in which the image total-variation (TV) (i.e., the ℓ_1-norm of the gradient magnitude image) is minimized, subject to constraints on data fidelity and image positivity.

$$\mathbf{f}^* = \operatorname{argmin}||\mathbf{f}||_{\mathrm{TV}} \quad \text{s.t.} \quad D(\mathbf{f}) \leq \epsilon \text{ and } f_n \geq 0, \tag{2}$$

where $D(\mathbf{f})$ denotes Euclidean data divergence between the measured data and model data:

$$D(\mathbf{f}) = |\mathcal{H}\mathbf{f} - \mathbf{g}|, \tag{3}$$

$||\mathbf{f}||_{\mathrm{TV}}$ the image TV. Parameter $\epsilon > 0$ determines a level of allowable inconsistency between the measured data and model data. We have also derived a necessary condition $c_\alpha(\mathbf{f}) = -1$ on algorithm convergence (Details and physical meaning of this parameter can be found in Refs. [10], [11]). For real data studies, our experience is that there is no significant change in the images when $c_\alpha(\mathbf{f})$ is less than -0.5.

We used a modified adaptive steepest descent and projection-onto-the-convex-set (ASD-POCS) algorithm [10], [11] for solving the optimization program. The pseudocode of ASD-POCS algorithm can be found in Refs. [9], [10]. The modification to the ASD-POCS algorithm is that: When $D(\mathbf{f}) \leq \epsilon$ is achieved, we then use a gradient descent method, instead of the POCS method, for further calculating $D(\mathbf{f})$ until $c_\alpha(\mathbf{f})$ is small enough [10], [11]. In this work, an initial image $\mathbf{f} = 0$ is used in the ASD-POCS reconstruction.

C. Reconstruction characterization

A considerable number of patient-data sets have been collected during the clinical trial of the breast-CT scanner, including cases with different breast sizes and densities. The scanning parameters are chosen according to the sizes and densities of the breasts. We then reconstructed images using the algorithms described above from these patient cases and compared the images reconstructed by use of ASD-POCS algorithms to those images currently reconstructed by use of FDK algorithm. The general procedures for selection of algorithm parameters has been described in detail in Refs. [16], [17]. Visual inspections are first performed to compare the two sets of reconstruction images for each patient case. We have also computed image power spectra for characterizing anatomical noise from images reconstructed by using ASD-POCS algorithm and FDK algorithm. According to [18]–[20], the power spectra of breast CT images can be characterized by a power law: $S(k) = \alpha/k^\beta$, where k is the spatial frequency. α and power-law exponent β are free parameters. Burgess and his colleagues [18] have shown that the power-law exponent is a critical factor in determining the size at which a lesion reaches detection threshold and a lower power-law exponent results in earlier detection of a growing lesion. Recognizing that the image power spectra does not strictly obey assumptions of shift invariance and noise stationarity, it still provides a useful, quantitative, relative metric of the noise magnitude and the power-law exponent can still provide certain measurement of the anatomical noise in breast images. We also performed a preliminary human-observer study in which the observers were asked their preference between images reconstructed by use of FDK and the proposed algorithms and will report the results at the conference.

III. Results

We have reconstructed images using the algorithms described above from a considerable number of patient-data sets of different breast sizes and densities scanned with different parameters. The variety of the patient cases provides a thorough check for the robustness of the reconstruction algorithms. In our reconstructions, effort was made to minimize the blocky artifacts that are typically observed in images reconstructed by use of TV-minimization-based algorithms from highly noisy data sets.

A. Visualization-based reconstruction characterization

We display in Figs. 1, 2, and 3 images reconstructed from three selected cases but plan to report additional cases at the

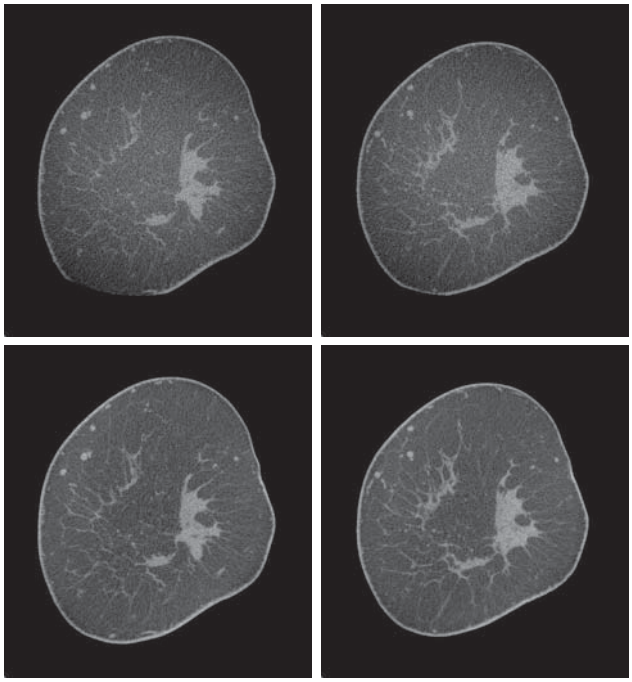


Figure 1. Images currently reconstructed by use of the FDK algorithm (top) and image reconstructed by use of the ASD-POCS algorithm (bottom) for patient case 1 at different distances to the middle plane. The display window is $[0.15, 0.25]$ cm^{-1}.

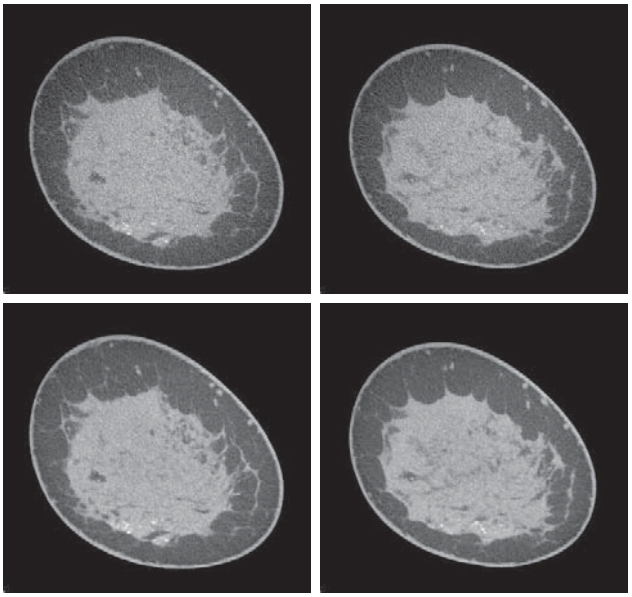


Figure 2. Image currently reconstructed by use of the FDK algorithm (top) and image reconstructed by use of the ASD-POCS algorithm (bottom) for patient case 2 at different distances to the middle plane. The display window is $[0.15, 0.25]$ cm^{-1}.

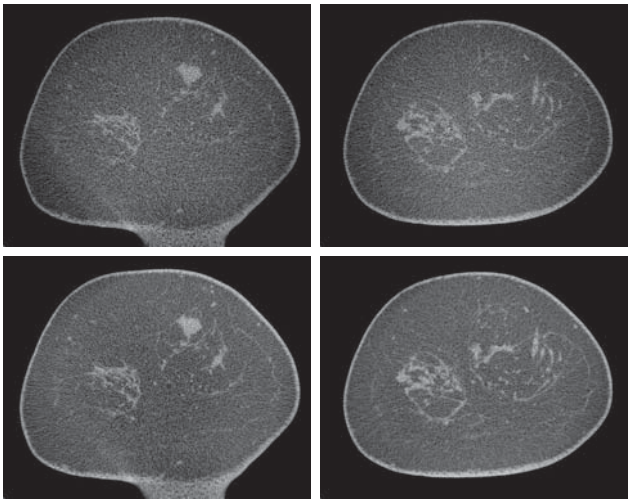


Figure 3. Image currently reconstructed by use of the FDK algorithm (top) and image reconstructed by use of the ASD-POCS algorithm (bottom) for patient case 3 at different distances to the middle plane. The display window is $[0.15, 0.25]$ cm^{-1}.

conference. For each case, we display slices at different distances to the middle plane. The images currently reconstructed by use of FDK algorithm are displayed in the top row and the corresponding slices reconstructed by use of ASD-POCS algorithms are displayed in the bottom row. Images of patient cases 1 are displayed in Fig. 1, which has fine structures in the glandular tissues. Upon careful examination, it can be observed that images reconstructed by use of ASD-POCS algorithm show enhanced details and better contrasts. We also display breast images of patient case 2 in Fig. 2, which has calcifications. It can be observed that the two sets of images are overall comparable. However, the small calcifications can be better discerned in the images reconstructed by use of ASD-POCS algorithms. In Fig. 3, we display a patient case with large breast but scanned with a small mAs, which means the projection data are of lower SNR than other cases. It can be observed that the reconstruction images are also noisier than those of the other two cases. However, the fine details in images reconstructed by use of ASD-POCS algorithms are still more enhanced than images currently reconstructed by use of FDK algorithms, which is similar to other cases.

B. Image-power-spectra-based characterization

From the reconstruction images, we also calculated image power spectra [19], [21] for characterizing anatomical noise of the breasts images. In order to calculate the image fluctuations in a small ROI within a reconstructed image, we select L square-shaped ROIs of equal size at different locations within the breast. Each of the ROIs is within the breast skin, and each ROI can overlap with other ROI but no more than half of its size. The image-power spectrum is first calculated on a Cartesian grid. We then converted power spectra onto a polar grid, followed by averaging them over the polar angle to obtain power spectra $P_p(k)$ as functions only of discrete radial frequencies k [11], [19]. For each case, we use slices within middle plane for the calculation.

We calculated power spectra from images reconstructed by using FDK algorithm and ASD-POCS algorithm at different iterations. In Fig. 4, we plot the power spectra calculated from images reconstructed by using ASD-POCS algorithms together with power spectra calculated from images currently reconstructed by use of FDK for two patient cases. It can be observed that power spectra of FDK images are in between the power spectra of ASD-POCS images at 10th and 100th iterations for both cases. We have also performed curve fitting and estimated the power-law exponent β. The β values corresponding to the plot on the left of Fig. 4 are 2.92 (FDK), 2.57 (ASD-POCS, 100th iteration), and 3.14 (ASD-POCS, 10th iteration). And the

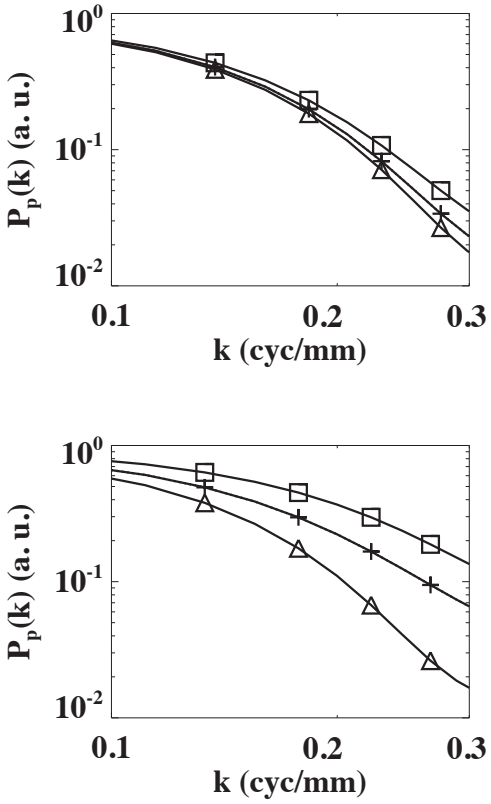


Figure 4. Image power spectra calculated from patient breast images reconstructed by use of FDK algorithms (+) and ASD-POCS algorithms at 10th (△) and 100th (□) iterations.

β values corresponding to the plot on the right of Fig. 4 are 1.99 (FDK), 1.44 (ASD-POCS, 100th iteration), and 3.13 (ASD-POCS, 10th iteration). It can be seen that, for both cases, images reconstructed by use of ASD-POCS algorithm have a smaller β. We have also calculated power spectra from images at other iterations reconstructed by use of ASD-POCS algorithm and the power spectra show little changes after around 50-80 iterations. We also calculated power spectra for other patient cases and similar results have be obtained.

IV. Discussion

In the work, we have investigated image reconstruction from low-SNR patient breast data in breast CT studies. The clinical data sets are of different breast sizes, densities, and different scanning protocols, and they provide a rigorous test for the robustness of the reconstruction algorithms. Based upon visual inspection, the TV-minimization-based reconstruction seems to reveal more clearly detailed structures with enhanced contrast than the FDK algorithm. We also investigated the power spectra of the reconstruction. Burgess and his colleagues [18] showed that the power-law exponent β has relationship with the pre-whitening observer performance measured by the contrast-detail diagram. For lesions of same size, a smaller β would result in a lower contrast limit for detecting the lesions by an ideal observer. Thus, if the scanning protocols of breast CT remain unchanged, this may imply that a low-contrast lesion in the images reconstructed by use of ASD-POCS algorithm can be better identified than those in the images currently reconstructed by use of FDK algorithms. This fact may be utilized either to improve the lesion detection or further reduce imaging dose for the current breast CT.

References

[1] C. H. J. Chang, J. L. Sibala, S. L. Fritz, S. J. Dwyer III, and A. W. Templeton, "Specific value of computed tomographic breast scanner (CT/M) in diagnosis of breast diseases," *Radiology*, vol. 132, pp. 647–652, 1979.

[2] J. J. Gisvold, D. F. Reese, and P. R. Karsell, "Computed tomographic mammography (CTM)," *Am. J. Roentgenol.*, vol. 133, pp. 1143–1149, 1979.

[3] C. H. J. Chang, D. E. Nesbit, D. R. Fisher, S. L. Fritz, S. J. Dwyer III, A. W. Templeton, F. Lin, and W. R. Jewell, "Computed tomographic mammography using a conventional body scanner," *Am. J. Roentgenol.*, vol. 138, pp. 553–558, 1982.

[4] J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated breast CT: Radiation dose and image quality evaluation," *Radiology*, vol. 221, pp. 657–667, 2001.

[5] B. Chen and R. Ning, "Cone-beam volume CT breast imaging: Feasibility study," *Med. Phys.*, vol. 29, pp. 755–770, 2002.

[6] J. M. Boone, N. Shah, and T. R. Nelson, "A comprehensive analysis of DgN(CT) coefficients for pendant-geometry cone-beam breast computed tomography," *Med. Phys.*, vol. 31, pp. 226–235, 2004.

[7] C. N. Brzymialkiewicz, M. P. Tornai, R. L. McKinley, and J. E. Bowsher, "Evaluation of fully 3-d emission mammotomography with a compact cadmium zinc telluride detector," *IEEE Trans. Med. Imag.*, vol. 24, pp. 868–877, 2005.

[8] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, pp. 612–619, 1984.

[9] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. and Technol.*, vol. 14, pp. 119–139, 2006.

[10] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008, PMCID:PMC2630711.

[11] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT ," *Phys. Med. Biol.*, vol. 55, pp. 6575–6599, 2010, PMCID: PMC3597413.

[12] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, E. L. Ritman, and X. Pan, "Algorithm-enabled low-dose micro-CT imaging," *IEEE Trans. Med. Imag.*, vol. 30, pp. 606–620, 2011, NIHMSID: NIHMS442722 (in process).

[13] M. Defrise, C. Vanhove, and X. Liu, "An algorithm for total variation regularization in high-dimensional linear problems," *Inverse Probl.*, vol. 27, p. 065002, 2011.

[14] E. A. Rashed and H. Kudo, "Statistical image reconstruction from limited projection data with intensity priors," *Phys. Med. Biol.*, vol. 57, pp. 2039–2061, 2012.

[15] P. T. Lauzier, J. Tang, and G.-H. Chen, "Prior image constrained compressed sensing: Implementation and performance evaluation," *Med. Phys.*, vol. 39, pp. 66–80, 2012.

[16] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan, "Optimization-based reconstruction of sparse images from few-view projections," *Phys. Med. Biol.*, vol. 57, pp. 5245–5273, 2012, PMCID:PMC3446871.

[17] J. Bian, J. Wang, X. Han, E. Y. Sidky, L. Shao, and X. Pan, "Optimization-based image reconstruction from sparse-view data in offset-detector CBCT," *Phys. Med. Biol.*, vol. 58, pp. 205–230, 2013, PMCID: PMC3590115.

[18] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Med. Phys.*, vol. 28, pp. 419–437, 2001.

[19] K. G. Metheany, C. K. Abbey, N. Packard, and J. M. Boone, "Characterizing anatomical variability in breast CT images," *Med. Phys.*, vol. 35, pp. 4685–4694, 2008.

[20] L. Chen, C. K. Abbey, and J. M. Boone, "Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities," *Phys. Med. Biol.*, vol. 58, pp. 1663–1681, 2013.

[21] G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," *Med. Phys.*, vol. 37, pp. 1948–1965, 2010.

Maximum *A Posteriori* Image Reconstruction for A High Sensitivity Uncollimated Small-Animal SPECT System

Jian Zhou, Katherine L. Walker, Gregory S. Mitchell, Simon R. Cherry, and Jinyi Qi

***Abstract*—A high sensitivity single-photon emission computed tomography (SPECT) system has been constructed by positioning two detectors in close proximity of the subject without using any collimator. In this paper, we present a reconstruction method for an uncollimated SPECT system recently reported in [1] and [2]. We derived an analytic formula for accurate calculation of the system matrix. It can model the solid angle effect, photon penetration and crystal gaps. We also developed a maximum *a posteriori* (MAP) reconstruction method to regularize the noise by using an image smoothness prior. We verified our analytic model using Monte Carlo simulations. Experimental results show the efficacy of the proposed MAP method.**

I. Introduction

High sensitivity preclinical single-photon emission computed tomography (SPECT) system has many advantages such as rapid imaging capability and low radiation dose. It has various applications in molecular imaging where good spatial resolution is not necessary. Mitchell and Cherry [1] have developed a prototype high-sensitivity SPECT imaging system with no collimator in Fig. 1. It consists of two planer, parallel scintillator detector heads. Such an uncollimated SPECT system has extremely high sensitivity owning to its compact geometry and wide coverage of solid angle. A 40% sensitivity has been reported for ^{99m}Tc with 3-mm thick NaI(Tl) scintillation detectors.

Tomographic image reconstruction for this uncollimated SPECT system is very challenging. The measured data are only two individual two-dimensional projection images while the image being reconstructed is three dimensional. Thus the reconstruction problem is severely underdetermined as compared to conventional tomography systems. In addition, because of no collimators the typical system response generates highly blurred measurement data that adds more degrees of ill-posedness to reconstruction. Mitchell and Cherry [1] used the maximum-likelihood expectation-maximization (ML-EM) algorithm and singular-value decomposition (SVD). Barrett and Holen [3] studied a similar uncollimated SPECT system which they referred to as a proximity-based imaging system. They described an analytic SVD reconstruction method based upon a pure solid angle-based system model. This model allowed them to derive an iterative Landweber algorithm that approaches reconstruction without explicitly calculating the system matrix.

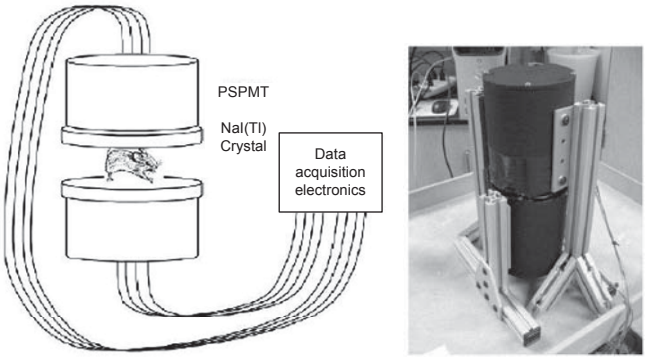


Fig. 1. A conceptual design of high-sensitivity preclinical SPECT system (left) and the first-generation prototype device (right). Pictures are taken from [1].

In this paper, we propose a maximum *a posteriori* (MAP) reconstruction method for a new uncollimated SPECT system. MAP uses an image prior to regularize the image reconstruction and to avoid the ill-posedness of the inverse problem. Rather than using the EM algorithm, we developed a preconditioned conjugate gradient (PCG) algorithm to maximize the log-posterior density function.

The remaining sections of this paper is organized as follows. In Section II, we describe the imaging model of the uncollimated SPECT system and MAP reconstruction. An analytic formula will be derived to calculate the system matrix. In Section III, we first validate our analytic system model by comparing it with Monte Carlo simulations. Then we apply our MAP method to experimental phantom data acquired from the second-generation uncollimated SPECT system [2]. Finally, we draw conclusions and discuss future work in Section IV.

II. Methods

A. System model

To model the physics of the data acquisition process, we start with a simple model ignoring scattering effects in the object and detector crystals. Let μ_c be the linear attenuation coefficient of the detector material for a given photon energy. For an infinitesimal volume $dV \equiv dudvdw$ inside this detector, the probability that a single photon emitted from a voxel

Department of Biomedical Engineering, University of California, Davis, CA 95616, USA. Email: {jnzhou, klbyrne, gsmitchell, srcherry, qi}@ucdavis.edu

This work is supported by the Office of Science (BER), U.S. Department of Energy under grand no DE-SC0005311.

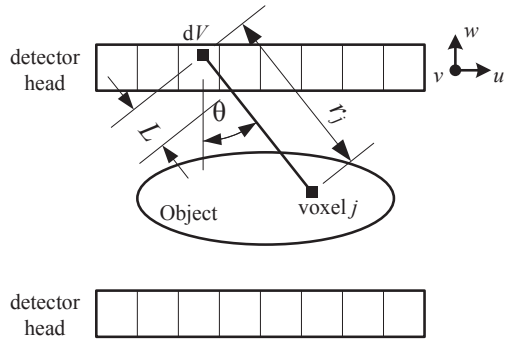


Fig. 2. An illustration of single photon detection.

j being detected by this volume is proportional to

$$\text{prob} \propto \left(\frac{\cos\theta du dv}{r_j^2}\right)\left(a_o e^{-\mu_c L}\right)\left(1-\exp\left(-\mu_c \frac{dw}{\cos\theta}\right)\right) \tag{1}$$

where θ is the angle between the photon travel path and the detector normal direction, r_j is the distance between voxel j and dV, L is the photon travel distance inside the detector before reaching dV. a_o represents the effect of object attenuation along the photon travel path. See Fig. 2 for an geometrical illustration.

The first term on the right side of (1) accounts for the solid angle effect based on the particular system geometry; the second term models the attenuation effect due to the object and the scintillator material itself, which describes the probability that a photon can reach dV; the third term is the probability that a photon can be absorbed inside dV. Using the approximation: $1-e^{-t} \approx t$ when t is small, (1) can be simplified to

$$\text{prob} \propto \frac{a_o \mu_c e^{-\mu_c L}}{r_j^2} dV. \tag{2}$$

Now considering a pixelated detector whose ith crystal is with a volume V_i, the probability of a photon emitted from voxel j being detected by crystal i can be calculated by

$$p_{ij} = \epsilon_i \mu_c \iiint_{V_i} \frac{a_o e^{-\mu_c L}}{r_j^2} dV \tag{3}$$

where we have included an additional factor ϵ_i that models the ith crystal efficiency. In practice, we compute this quantity using numerical integration. Note that a_o, L and r_j can vary when sampling each individual crystal. Also in (3) we have assumed that each image voxel behaves like a delta function. For other advanced voxel functions, additional volume integration in image space may be applied.

B. Image reconstruction

Let y_i be the number of photons detected by crystal i. Define $\boldsymbol{y} = [y_1, \ldots, y_I]^T$ (where T denotes the vector or matrix transpose) as a collection of measurements by a total number of I crystals. We assume that $\boldsymbol{y}$ is Poisson distributed whose mean, denoted by $\bar{\boldsymbol{y}}$, is

$$\bar{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{x} + \bar{\boldsymbol{s}} \tag{4}$$

where $\boldsymbol{x} = [x_1, \ldots, x_J]^T$ is the unknown image with a total number of J voxels, $\boldsymbol{P}$ is the system matrix whose element p_{ij} is determined by (3), and $\bar{\boldsymbol{s}} = [s_1, \ldots, s_I]^T$ models the mean of background scatters.

For a given $\boldsymbol{y}$, a MAP reconstruction of $\boldsymbol{x}$ is found as

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\geqslant 0} \Phi(\boldsymbol{x}), \quad \Phi(\boldsymbol{x}) = L(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x}) \tag{5}$$

where $L(\boldsymbol{y}|\boldsymbol{x})$ is the Poisson log-likelihood function

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{I} \{y_i \log(\bar{y}_i) - \bar{y}_i - \log(y_i!)\} \tag{6}$$

and $U(\boldsymbol{x})$ is the prior energy function, and β is a parameter adjusting the strength of the image prior. In this paper we use a Gaussian prior whose energy function takes the form of

$$U(\boldsymbol{x}) = \frac{1}{2}\sum_{j=1}^{J}\sum_{k\in\mathcal{N}_j} \omega_{jk}(x_k - x_j)^2 \tag{7}$$

where $\mathcal{N}_j$ represents the neighborhood of voxel j and ω_{jk} is a weighting factor that is chosen to be the inverse distance between voxels j and k.

Since the MAP reconstruction has no closed-form solution, an iterative algorithm has to be used. The classic EM algorithm has slow convergence, so here we adapt the PCG algorithm proposed in [4] which converges faster. An advantage of using an iterative method as compared to the SVD methods [1][3] is that we can enforce the nonnegativity constraint.

III. Results

A. Model verification

To verify the accuracy of our analytic system model, we simulated the uncollimated SPECT system in GATE [5]. It consists of two identical planar sodium iodide scintillation detectors separated by 22 mm (face to face). Each detector is 3-mm thick and contains a 50 (u) $\times$ 25 (v) crystal array with crystal pitch of 2 mm (u) $\times$ 2 mm (v). A spherical monoenergetic gamma-ray source of diameter 0.2 mm was placed at location (u = 0 mm, v = 0 mm, w = 5 mm) (see Fig. 3). We simulated three energies: 30 KeV, 140 KeV and 511 KeV. For analytic calculation, we forward projected each point source using a precomputed system matrix with no object attenuation, and no crystal efficiency variability and no background scatters.

Fig. 4–6 compares analytic calculated singles images and those obtained from Monte Carlo simulations. We see that

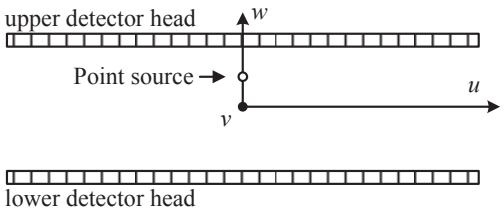


Fig. 3. A simulated uncollimated SPECT system geometry for point source acquisition.

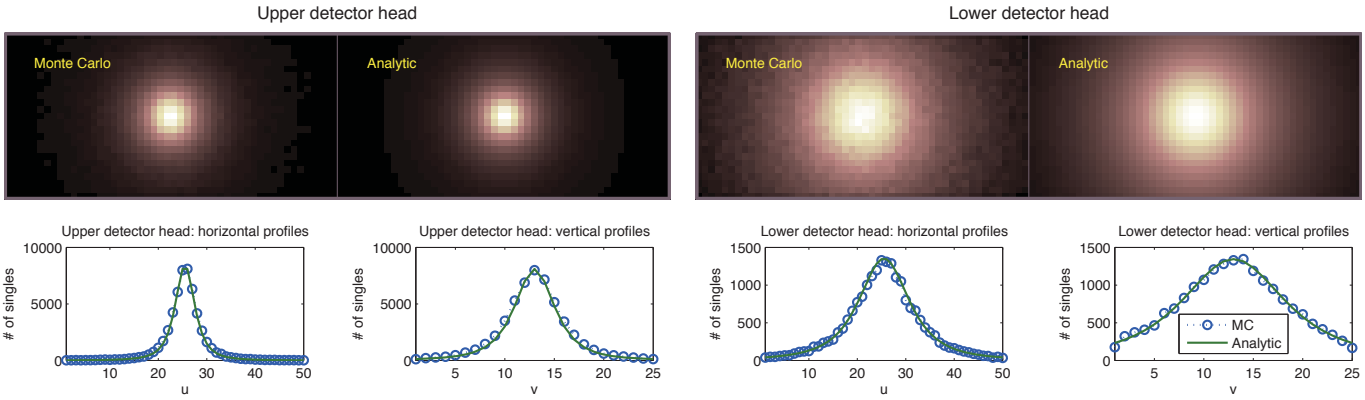


Fig. 4. A comparison for point source at 30 KeV. Profiles are taken across the center of image.

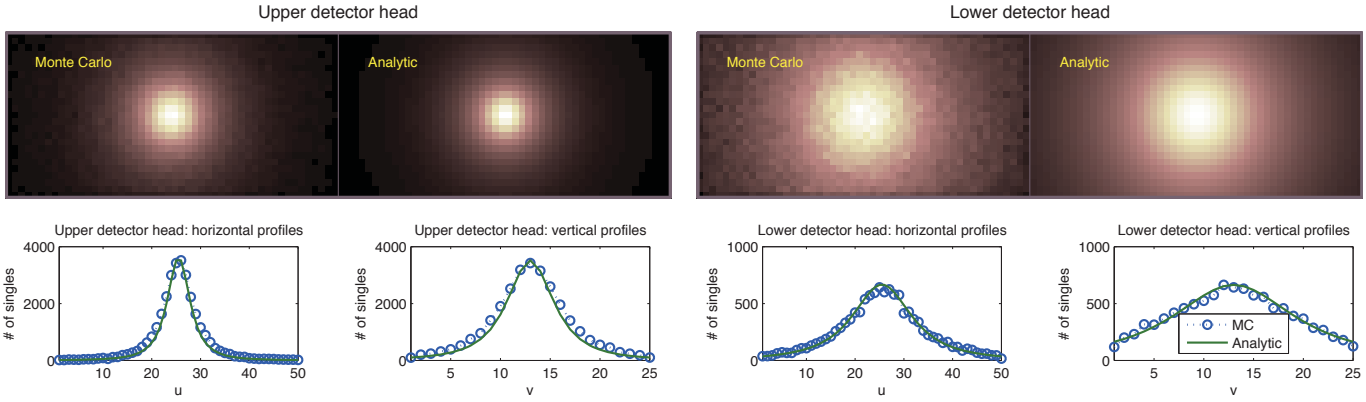


Fig. 5. A comparison for point source at 140 KeV. Profiles are taken across the center of image.

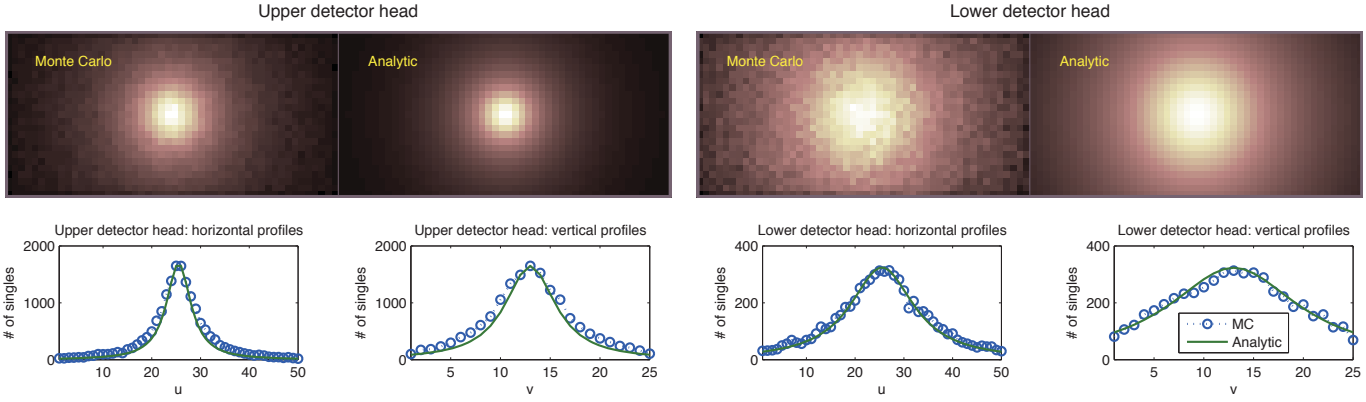


Fig. 6. A comparison for point source at 511 KeV. Profiles are taken across the center of image.

our calculation agrees very well with Monte Carlo results. This validates the proposed analytic formula (3). We observed slight mismatches in the profiles when the source energy is higher. We believe that it is caused by inter-crystal scatter effect which is not considered in our analytic model.

B. Experimental phantom study

We acquired real phantom data using the second-generation uncollimated SPECT system. Each detector head contains one scintillator crystal array read out by two Hamamatasu H8500 multichannel photomultiplier tubes (see Fig. 7). Details on the system design can be found in [2]. The two detectors are separated by 25 mm. In our experimental study, we used two pixelated NaI(Tl) arrays with different pixel dimensions: 1.5 mm (u) $\times$ 1.5 mm (v) $\times$ 3 mm (w) for the upper detector and 2.0 mm (u) $\times$ 2.0 mm (v) $\times$ 3 mm (w) for the lower detector. The gap between crystals is about 0.2 mm. The number of pixels are 41 (u) $\times$ 21 (v) and 51 (u) $\times$ 24 (v) for the upper and the lower detector, respectively. The total active areas of the two detectors are roughly the same.

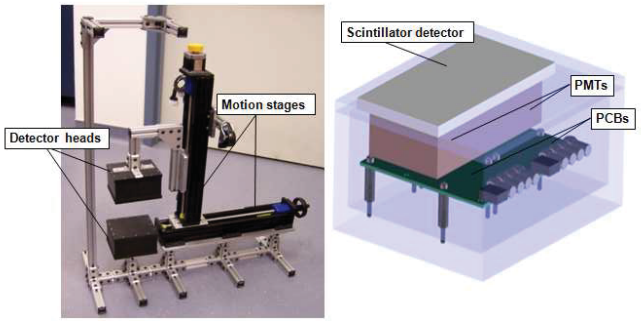
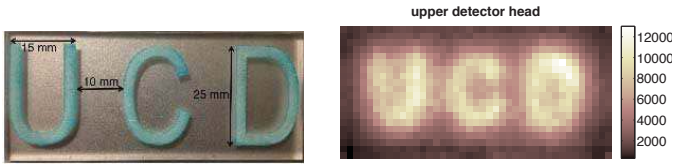


Fig. 7. The second-generation uncollimated SPECT imaging system. Picture is taken from [2].



(a) UCD phantom (b) Measured singles image

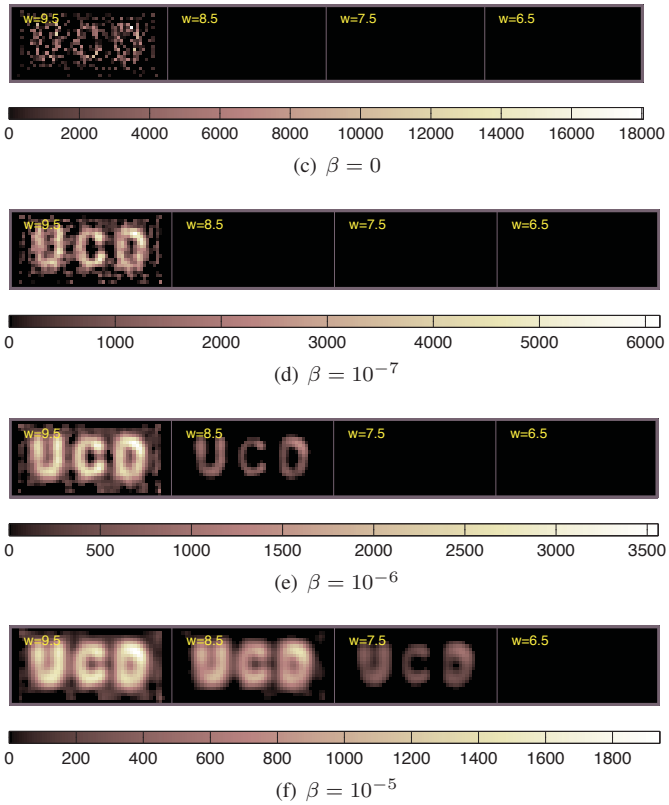


Fig. 8. A comparison of MAP reconstruction for a letter phantom. In (c)-(f), each subimage is an image plane perpenticular to the w axis. The distance between each plane to the origin is indicated in label (unit in mm).

A phantom consisting of three letters ('UCD') less than 1 mm deep was created and filled with ^{99m}Tc (energy 140 KeV) as shown in Fig. 8(a). The phantom was placed against the upper scintillation detector window (about 3 mm below the NaI(Tl) crystal). The total activity in the phantom was 100 nCi and a 10-min acquisition was performed. Fig. 8(b) shows the acquired singles image from the upper detector head. The total number of recorded singles counts is about 4M. Only a very loose energy cut was applied to the collected data, which had a trigger threshold of roughly 30 KeV. We precomputed a system matrix using an image dimension 51 (u) $\times$ 25 (v) $\times$ 20 (w) with a voxel size 2 mm (u) $\times$ 2 mm (v) $\times$ 1 mm (w). Neither object attenuation nor background scatters was considered. We ran 300 PCG iterations for all MAP reconstructions.

Fig. 8(c)-(f) show reconstructed images using various β values. Fig. 8(c) is equivalent to the ML reconstruction since β is equal to zero. Obviously it is very noisy. The MAP reconstructions shown in Fig. 8(d)-(f) are smoother, and one can easily recognize those three letters. The image quality varies with different β value. In this study, $\beta = 10^{-6}$ appears to be a good choice. Note that the spacing between letter 'U' and letter 'C' is 10 mm, so a spatial resolution less than 10 mm (in u and v directions) is clearly achievable near the detector head.

IV. Conclusions

We have studied image reconstruction for a uncollimated preclinical SPECT imaging system. We have derived an analytic model for system matrix calculation, which achieved good accuracy for low energy photons. To overcome the ill-posedness of image reconstruction, we have developed an MAP reconstruction with a Gaussian image prior. Our preliminary results showed that the proposed MAP method works very well with real experimental data. Our ongoing work includes a study on edge-preserved image priors. In addition, more experimental studies using small animals and plants will be performed in future.

References

[1] G. S. Mitchell and S. R. Cherry, "A high-sensitivity small animal SPECT system," *Physics in Medicine and Biology*, vol. 54, no. 5, pp. 1291–1305, 2009.

[2] K. L. Byrne, J. Zhou, J. Qi, S. R. Cherry, and G. S. Mitchell, "Characterization of detectors for a high sensitivity SPECT system for small animals and plants," in *Nuclear Science Symposium, 2012. IEEE, Poster presentation M21-34*, Nov 2012.

[3] H. H. Barrett and R. Van Holen, "Analytical singular-value decomposition of three-dimensional, proximity-based SPECT systems," in *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2011, pp. 100–103.

[4] J. Qi, R. Leahy, C. Hsu, T. Farquhar, and S. Cherry, "Fully 3D bayesian image reconstruction for the ECAT EXACT HR+," in *Nuclear Science Symposium, 1997. IEEE*, vol. 2, Nov 1997, pp. 953 –957 vol.2.

[5] S. Jan and *et al*, "Gate: a simulation toolkit for PET and SPECT," *Physics in Medicine and Biology*, vol. 49, no. 19, pp. 4543–4561, 2004.

Analytic methods for list-mode reconstruction

Abhinav K. Jha, Harrison H. Barrett, Eric Clarkson, Luca Caucci, and Matthew A. Kupinski

***Abstract*—List-mode (LM) acquisition of imaging-system data does not suffer from information loss due to data binning. To take advantage of this acquisition mechanism, efficient methods are required to perform object reconstruction using LM data. Current methods to perform reconstruction using LM data reconstruct discrete representation of the object, but since object functions are essentially defined on continuous domains, this leads to information loss. In this paper, we exploit the fact that LM data are defined on a continuous domain, and design analytic methods to reconstruct the object function from LM data. A general procedure to design analytic LM reconstruction algorithms is first formulated. We use this procedure to reconstruct the object function for a linear shift-invariant imaging system with a Gaussian point spread function. We then consider the problem of LM reconstruction in single-photon emission computed tomography (SPECT) imaging systems. We present an analytic method to perform LM reconstruction for a hypothetical SPECT system with an infinite object support and infinite angular sampling. We extend this method to finite angular sampling, and realize that due to the infinite support of the object, the reconstruction cannot be performed using our scheme, but with a finite support, such a reconstruction should be possible. The developed reconstruction schemes can aid in accurate comparison of LM and binned-data acquisition techniques from a task-based perspective.**

I. Introduction

List-mode (LM) acquisition and processing of data is gaining wide popularity for photon-counting imaging systems [1]–[4]. A major advantage of LM acquisition is that they do not suffer from information loss due to binning unlike the more conventional sinogram-based storage and processing of data. However, while LM data contain more information, in the absence of efficient information-retrieval algorithms, this extra information is not of much use. To retrieve this extra information from LM data, often the first step is to design methods to reconstruct the object from the LM data. Algorithms have been developed to reconstruct the object from LM data [1]–[3], but these methods reconstruct discrete representations of the object. The objects in imaging are functions defined on a continuous domain, and reconstructing a discrete representation of the object leads to information loss. A more appropriate methodology is to reconstruct the object as a function defined on a continuous domain. Our primary interest is in performing this object reconstruction for single-photon emission computed tomography (SPECT) imaging systems, but we will keep the problem general.

Interestingly, much literature on reconstruction in SPECT is based on a continuous-continuous (CC) formulation of the SPECT imaging system [5]. These analytic algorithms, such as the well-known analytical filtered-backprojection (FBP) algorithm, reconstruct the object function. However, most SPECT imaging systems bin the data, and therefore, the reconstruction methods developed for CC systems are modified to instead work with discrete image data, and reconstruct discrete object representations [5]–[7]. In this context, LM acquisition presents us with another advantage: It yields data that is defined on a continuous domain, and thus fits the CC formulation of the SPECT imaging system. The primary objective of this work is to design analytic algorithms that can exploit this advantage to reconstruct object functions from LM data defined on a continuous domain.

There are many motivations to design analytic reconstruction algorithms. Analytic algorithms can leverage the true potential of LM data by exploiting the CC nature of LM acquisition, and avoid information loss that would otherwise occur in algorithms that reconstruct discrete object representations. Another advantage of analytic algorithms is that they offer a method to compare information-retrieval techniques without being affected by the limitations of simulation studies, such as discretization requirement and floating-point issues. Analytical algorithms also offer insights on the information content and information-retrieval capacity from the data, which can help improve the design of the imaging system and the algorithm. There are also computational and economic advantages to using analytic approaches [5]. For example, although nonlinear iterative reconstruction algorithms can account for factors such as noise, they require significantly higher computation compared to the one-step analytic methods. Various assessment schemes can compare analytic algorithms with other reconstruction approaches, but these assessment schemes also require development of the analytic method.

We begin with deriving a general framework to reconstruct the object in any imaging system that acquires LM data, and then apply this framework to specific imaging systems.

II. General reconstruction approach

We assume that the object being imaged is a scalar-valued function of spatial position $\boldsymbol{r}$, where $\boldsymbol{r}$ is a vector with s components lying in $\mathbb{R}^s$. We will denote the object by the function $f(\boldsymbol{r})$ and assume that the object function lies in the Hilbert space $\mathbb{L}_2(\mathbb{R}^s)$. This object is viewed over a measurement time τ by some photon-counting imaging system, which detects the photons and then, for each detected photon, estimates attributes such as direction, energy, and the position of interaction of the photon with the detector. For the j^{th} event, these estimated attributes are grouped into a q-dimensional (q-D) vector $\hat{\boldsymbol{A}}_j$. The LM data can be described using the point

A. K. Jha, H. H. Barrett, E. Clarkson, and M. A. Kupinski are with the College of Optical Sciences, University of Arizona, Tucson AZ, USA.

H. H. Barrett, E. Clarkson, L. Caucci, and M. A. Kupinski are with the Department of Medical Imaging, University of Arizona, Tucson AZ, USA.

This work was supported by National Institute of Biomedical Imaging and Bioengineering of National Institute of Health under grant number RC1-EB010974, R37-EB000803 and P41-EB002035.

process $u(\hat{\boldsymbol{A}})$ given by

$$u(\hat{\boldsymbol{A}}) = \sum_{j=1}^{J} \delta(\hat{\boldsymbol{A}} - \hat{\boldsymbol{A}}_j), \tag{1}$$

where $\delta(\ldots)$ denotes the Dirac delta function. Taking the mean of this point process gives [4]

$$\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau) = \int_{\mathbb{S}_f} d^s r \ \tau \mathrm{pr}(\hat{\boldsymbol{A}}|\boldsymbol{r}) s(\boldsymbol{r}) f(\boldsymbol{r}), \tag{2}$$

where $s(\boldsymbol{r})$ denotes the sensitivity of the detector to activity occurring at location $\boldsymbol{r}$ and $\mathbb{S}_f$ denotes the support of the functions in object space. Eq. (2) can be written in operator form as

$$\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau) = [\mathcal{L}\boldsymbol{f}](\hat{\boldsymbol{A}}). \tag{3}$$

where $\mathcal{L}$ denotes the linear LM imaging operator. The kernel for the operator $\mathcal{L}$ is given by

$$l(\hat{\boldsymbol{A}},\boldsymbol{r}) = \tau \mathrm{pr}(\hat{\boldsymbol{A}}|\boldsymbol{r}) s(\boldsymbol{r}). \tag{4}$$

In LM acquisition, the function $\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau)$ lies in $\mathbb{L}_2(\mathbb{R}^{q-t}) \times \mathbb{E}^t$, where $\mathbb{E}^t$ denotes the t-D Eucledian space, and where $q > t$. Thus, the operator $\mathcal{L}$ maps from the set of functions $f(\boldsymbol{r})$ that lie in $\mathbb{L}_2(\mathbb{R}^s)$ to the set of functions that lie in $\mathbb{L}_2(\mathbb{R}^{q-t}) \times \mathbb{E}^t$. If $\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau)$ lies in $\mathbb{L}_2(\mathbb{R}^q)$, i.e. if $t = 0$, and if $q \geq s$, then it is possible that the operator $\mathcal{L}$ has no null space. This is unlike in binned-data acquisition, where we map from the set of functions lying in $\mathbb{L}_2(\mathbb{R}^s)$ to vectors in the $\mathbb{E}^q$, a mapping that definitely has null space. Even if $\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau)$ lies in $\mathbb{L}_2(\mathbb{R}^{q-t}) \times \mathbb{E}^t$ where $q > t$, the operator $\mathcal{L}$ should have a smaller null space compared to a system in which data is binned. Since the operator $\mathcal{L}$ might have no null space or a reduced null space, LM acquisition provides an avenue to reconstruct the object with lesser information loss compared to binned-data acquisition.

To derive the reconstruction technique, we determine the expression for the pseudoinverse of the LM operator, which requires performing a singular value decomposition (SVD) of the $\mathcal{L}$ operator. Let us denote the singular values and the singular vectors of $\mathcal{L}$ in object and data space by $\mu_i, \boldsymbol{w}_i$, and $\boldsymbol{v}_i$, respectively. The pseudoinverse of the $\mathcal{L}$ operator, which we denote by $\mathcal{L}^+$, can be represented as

$$\mathcal{L}^+ = \sum_{i=1}^{R} \frac{1}{\sqrt{\mu_i}} \boldsymbol{w}_i \boldsymbol{v}_i^\dagger. \tag{5}$$

Our reconstruction approach is to apply this pseudoinverse to the acquired noisy image data:

$$\hat{f}(\boldsymbol{r}) = [\mathcal{L}^+ u](\boldsymbol{r}). \tag{6}$$

The reconstruction in the noise-free case is given by $\hat{f}_{\mathrm{nfree}}(\boldsymbol{r}) = [\mathcal{L}^+ \bar{u}](\boldsymbol{r})$. We can verify that Eq. (6) will lead to this solution in a mean sense, i.e. $\langle \hat{f}(\boldsymbol{r}) \rangle_{u|f} = \hat{f}_{\mathrm{nfree}}(\boldsymbol{r})$, where $\langle \cdots \rangle$ denotes the mean of the quantity inside the parenthesis.

To compute the SVD of $\mathcal{L}$, we must determine the expression for $\mathcal{L}^\dagger\mathcal{L}$, where $\mathcal{L}^\dagger$ denotes the adjoint of $\mathcal{L}$. We can derive the kernel for the $\mathcal{L}^\dagger\mathcal{L}$ operator, which we denote by $k(\boldsymbol{r},\boldsymbol{r}')$ to be

$$k(\boldsymbol{r}',\boldsymbol{r}) = \tau^2 \int_{\mathbb{S}_u} d^q \hat{A} \ s(\boldsymbol{r}) s(\boldsymbol{r}') \mathrm{pr}(\hat{\boldsymbol{A}}|\boldsymbol{r}) \mathrm{pr}(\hat{\boldsymbol{A}}|\boldsymbol{r}'), \tag{7}$$

where $\mathbb{S}_u$ denotes the LM-data support. We now investigate whether, for specific imaging systems, we can perform the SVD of the $\mathcal{L}$ operator.

III. A system with Gaussian point spread function

Let us consider a simple imaging system that satisfies similar assumptions as made in Caucci et al. [4]. The system consists of a 2-D object $f(\boldsymbol{r})$ imaged to a 2-D detector. The LM attributes acquired are the x and y coodinates of the detection. Thus $\hat{\boldsymbol{A}}$ is a 2-D vector, which we henceforth denote by $\hat{\boldsymbol{R}}$. The optics of the imaging system and the detector is assumed to be linear and shift invariant (LSIV), and characterized by Gaussian point spread functions (PSFs). With these assumptions, we can obtain that $\hat{\boldsymbol{R}}$ conditioned on $\boldsymbol{r}$ is normally distributed:

$$\mathrm{pr}(\hat{\boldsymbol{R}}|\boldsymbol{r}) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(\hat{\boldsymbol{R}} - \boldsymbol{r})^2}{2\sigma^2}\right], \tag{8}$$

where σ^2 is the sum of the variances due to the optics of the imaging system and the detector. We also assume that $s(\boldsymbol{r})$ is equal to unity for all $\boldsymbol{r}$. Using Eqs. (4) and (8), we find that

$$l(\hat{\boldsymbol{R}},\boldsymbol{r}) = \frac{\tau}{2\pi\sigma^2} \exp\left[-\frac{(\hat{\boldsymbol{R}} - \boldsymbol{r})^2}{2\sigma^2}\right], \tag{9}$$

and thus $\mathcal{L}$ resembles a convolution operator. Therefore, the pseudoinverse of $\mathcal{L}$ can be derived to be represented as

$$\mathcal{L}^+ = \frac{1}{\tau} \int d^2\rho \ \exp(2\pi^2\sigma^2\rho^2) \exp\{2\pi i \boldsymbol{\rho} \cdot (\boldsymbol{r} - \hat{\boldsymbol{R}})\}. \tag{10}$$

where $\rho = |\boldsymbol{\rho}|$. Thus, for this imaging system, the object function can be reconstructed by taking the Fourier transform of the acquired data, dividing it by the Fourier transform of kernel of the LM operator at that frequency, and then taking the inverse Fourier transform of the result. In theory, the $\mathcal{L}^+$ operator should be applied to $\bar{u}(\hat{\boldsymbol{A}}|\boldsymbol{f},\tau)$, but in practice, it can be used as in Eq. (6) to obtain $\hat{f}(\boldsymbol{r})$ from $u(\hat{\boldsymbol{A}})$.

IV. Tomographic imaging system

Consider a tomographic 2-D SPECT imaging system in the x-y plane. The SPECT imaging system consists of a parallel-hole collimator with bores on a regular grid, followed by a 1-D detector. The system rotates about the z axis to acquire data at multiple angles θ, where θ denotes the angle that the detector makes with the x-axis. Let us assume that the parallel-hole collimator accepts photons only normal to the detector surface. Also, let us ignore attenuation in this analysis. In each LM event, the coordinate of the position of interaction of the gamma ray photon with the scintillation crystal is estimated and recorded. We denote the estimated position of interaction by $\hat{p}$, and its corresponding true value by p. The angular orientation of the detector θ is also recorded. To derive the expression for the $\mathcal{L}$ operator for this system, we have to determine the expression for $\mathrm{pr}(\hat{p},\theta|\boldsymbol{r})$. Using marginal probabilities, $\mathrm{pr}(\hat{p},\theta|\boldsymbol{r})$ can be written as

$$\mathrm{pr}(\hat{p},\theta|\boldsymbol{r}) = \mathrm{pr}(\theta|\boldsymbol{r}) \int dp \ \mathrm{pr}(\hat{p}|p,\theta,\boldsymbol{r}) \mathrm{pr}(p|\theta,\boldsymbol{r}). \tag{11}$$

The probability of the position of interaction p given a particular value of the detector angle θ and object location $\boldsymbol{r}$ is simply the delta function $\delta(p - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\theta)$, where $\hat{\boldsymbol{n}}_\theta$ is the normal to the detector face when the detector is aligned at an angle θ. Thus

$$\mathrm{pr}(p|\theta, \boldsymbol{r}) = \delta(p - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\theta). \tag{12}$$

Let us assume that $\hat{p}$ was estimated using a maximum-likelihood (ML) scheme, where all the scintillation photons were used to estimate the attribute. Then using the asymptotic properties of ML estimates, it can be shown that $\mathrm{pr}(\hat{p}|p, \theta, \boldsymbol{r})$ is normally distributed with the mean given by the true value p and the variance σ_p^2 given by the Cramér-Rao lower bound for the estimate on p [4]:

$$\mathrm{pr}(\hat{p}|p, \theta, \boldsymbol{r}) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(\hat{p} - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\theta)^2}{2\sigma_p^2}\right]. \tag{13}$$

Under the assumption that $s(\boldsymbol{r})$ is unity for all values of $\boldsymbol{r}$, using Eqs. (4), (11)-(13), we can obtain the kernel of the LM operator $\mathcal{L}$ to be

$$l(\hat{p}, \theta, \boldsymbol{r}) = \tau \mathrm{pr}(\theta|\boldsymbol{r}) \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(\hat{p} - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\theta)^2}{2\sigma_p^2}\right]. \tag{14}$$

Having derived the general form for the kernel of the LM operator, we now analyze the possibility of the pseudoinverse of the LM operator for some specific cases.

A. Infinite angular sampling and infinite object support

Consider a SPECT imaging system with infinite object support and infinite angular sampling. Due to the isotropic emission of photons, for this system

$$\mathrm{pr}(\theta|\boldsymbol{r}) = \frac{1}{2\pi}. \tag{15}$$

Inserting this expression into Eq. (14), the kernel of the LM operator is given by

$$l(\hat{p}, \theta, \boldsymbol{r}) = \frac{\tau}{2\pi} \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{(\hat{p} - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_\theta)^2}{2\sigma_p^2}\right]. \tag{16}$$

Using Eq. (7), the kernel $k(\boldsymbol{r}', \boldsymbol{r})$ for the $\mathcal{L}^\dagger\mathcal{L}$ operator can be derived to be

$$k(\boldsymbol{r}', \boldsymbol{r}) = \left[\frac{\tau}{2\pi}\right]^2 \frac{1}{2\sqrt{\pi}\sigma_p} \int d\theta \ \exp\left[\frac{-\{(\boldsymbol{r} - \boldsymbol{r}') \cdot \hat{\boldsymbol{n}}_\theta\}^2}{4\sigma_p^2}\right]. \tag{17}$$

We note that $k(\boldsymbol{r}', \boldsymbol{r})$ is a function of $\boldsymbol{r} - \boldsymbol{r}'$ and therefore, the eigenanalysis of the $\mathcal{L}^\dagger\mathcal{L}$ operator can be performed via Fourier analysis. The eigenvectors of $\mathcal{L}^\dagger\mathcal{L}$ are the complex exponentials given by

$$w(\boldsymbol{\rho}_0)(\boldsymbol{r}) = \exp(2\pi i \boldsymbol{\rho}_0 \cdot \boldsymbol{r}). \tag{18}$$

The corresponding eigenvalues for these eigenvectors are determined by computing the Fourier transform of the convolution kernel (Eq. (17)). Denoting the Fourier transform of this kernel at frequency $\boldsymbol{\rho}$ by $K(\boldsymbol{\rho})$, we can derive that

$$K(\boldsymbol{\rho}) = \frac{1}{\rho}\left[\frac{\tau}{2\pi}\right]^2 \exp(-4\pi^2\sigma_p^2\rho^2). \tag{19}$$

Expressing the vectors $\boldsymbol{r}$ and $\boldsymbol{\rho}$ in terms of the basis vectors $\hat{\boldsymbol{n}}_\theta$ and $\hat{\boldsymbol{n}}_{\perp,\theta}$ as $\boldsymbol{r} = r_1(\theta)\hat{\boldsymbol{n}}_\theta + r_2(\theta)\hat{\boldsymbol{n}}_{\perp,\theta}$ and $\boldsymbol{\rho} = \rho_1(\theta)\hat{\boldsymbol{n}}_\theta + \rho_2(\theta)\hat{\boldsymbol{n}}_{\perp,\theta}$, the singular vectors for the $\mathcal{L}$ operator in data space can be derived to be

$$v(\hat{p}, \theta) = \sqrt{\rho}\exp(-2\pi i \hat{p}\rho_1(\theta))\delta(\rho_2(\theta)). \tag{20}$$

Using Eqs. (5), (18)-(20) the pseudoinverse of the $\mathcal{L}$ operator is represented as

$$\mathcal{L}^+ = \frac{2\pi}{\tau}\int d\rho_1(\theta)\rho_1 \exp[2\pi i \rho_1(\theta)(r_1(\theta) + \hat{p})]\exp(2\pi^2\sigma_p^2\rho_1^2). \tag{21}$$

Using Eq. (6), the reconstructed object $\hat{f}(\boldsymbol{r})$ is given by

$$\hat{f}(\boldsymbol{r}) = \frac{2\pi}{\tau}\int d\theta \int d\hat{p} \int d\rho_1(\theta)\rho_1(\theta)\times \exp[2\pi i \rho_1(\theta)(r_1(\theta) + \hat{p})]\exp(2\pi^2\sigma_p^2\rho_1^2)u(\hat{p}, \theta). \tag{22}$$

B. Finite angular sampling and infinite object support

We now consider a more conventional SPECT system that acquires data at multiple angles θ_j, where the index j varies from 1 to J. For this system, each LM event consists of the position estimate $\hat{p}$ and the detector angle index j. Thus this system maps from a set of functions that lie in the space $\mathbb{L}_2(\mathbb{R}^2)$ to a set of functions that lie in the space $\mathbb{L}_2(\mathbb{R}) \times \mathbb{E}$, and the LM operator for this system has the kernel given by $l(\hat{p}, j, \boldsymbol{r})$. To determine the expression for this kernel, we consider the general expression given by Eq. (14). The expression for $\mathrm{Pr}(\theta|\boldsymbol{r})$ in this case is equal to $\frac{1}{J}$. Thus the expression for the kernel of the LM operator is given by

$$l(\hat{p}, j, \boldsymbol{r}) = \frac{\tau}{\sqrt{2\pi}\sigma_p J} \exp\left[-\frac{(\hat{p} - \boldsymbol{r} \cdot \hat{\boldsymbol{n}}_j)^2}{2\sigma_p^2}\right], \tag{23}$$

where $\hat{\boldsymbol{n}}_j$ denotes the normal to the detector surface, when the detector is aligned at angle θ_j. We can derive the kernel of the $\mathcal{L}^\dagger\mathcal{L}$ operator to be

$$k(\boldsymbol{r}', \boldsymbol{r}) = \left[\frac{\tau}{J}\right]^2 \frac{1}{2\sqrt{\pi}\sigma_p} \sum_{j=1}^{J} \exp\left[-\frac{\{(\boldsymbol{r} - \boldsymbol{r}') \cdot \hat{\boldsymbol{n}}_j\}^2}{4\sigma_p^2}\right]. \tag{24}$$

We again note that $\mathcal{L}^\dagger\mathcal{L}$ resembles a convolution operator, so its eigenvectors are the complex exponentials. The eigenvalues corresponding to these eigenvectors are determined by taking the Fourier transform of the convolution kernel in Eq. (24). These eigenvalues are given by

$$K(\boldsymbol{\rho}) = \left[\frac{\tau}{J}\right]^2 \sum_{j=1}^{J} \exp(-4\pi^2\sigma_p^2(\boldsymbol{\rho} \cdot \hat{\boldsymbol{n}}_j)^2)\delta(\boldsymbol{\rho} \cdot \hat{\boldsymbol{n}}_{\perp,j}), \tag{25}$$

where we have expressed $\boldsymbol{r}$ and $\boldsymbol{\rho}$ in terms of the basis vectors $\hat{\boldsymbol{n}}_j$ and $\hat{\boldsymbol{n}}_{\perp,j}$ as previously. Due to the delta function, the sum over j exists only when $\boldsymbol{\rho}$ is parallel to $\hat{\boldsymbol{n}}_j$. The delta function complicates further analysis. For example, to find the singular vectors in data space or to determine the expression for $\mathcal{L}^\dagger$, we must divide by the square root of the eigenvalues $K(\boldsymbol{\rho})$, which requires taking the square root of the delta function. We can avoid this issue when determining the singular vectors

in data space by obtaining the SVD representation of $\mathcal{L}$ and then using that to compute the data space singular vectors. However, the issue cannot be avoided when computing the pseudoinverse. Thus, it seems unlikely that the pseudoinverse of $\mathcal{L}$ can be determined for this case. A physical interpretation of the absence of the pseudoinverse is the following: The finite angular sampling leads to a set of null functions. Also, the infinite support leads to an infinite number of solutions when $\boldsymbol{\rho}$ is parallel to $\hat{\boldsymbol{n}}_j$, and thus the absence of the pseudo-inverse. Having a finite support for the object might cause this problem to disappear. We consider this case now.

C. Finite support and finite angular sampling

Following a similar treatment that led to Eq. (23) but constraining the object support $\mathbb{S}_f$ to be finite, we obtain the transformation from the object to the image space as:

$$\bar{u}(\hat{p}, j) = \frac{\tau}{J} \frac{1}{\sqrt{2\pi}\sigma_p} \int_{\mathbb{S}_f} d^2r \exp\left[-\frac{(\hat{p} - \boldsymbol{r}\cdot\hat{\boldsymbol{n}}_j)^2}{2\sigma_p^2}\right] f(\boldsymbol{r}). \tag{26}$$

Fourier analysis of the $\mathcal{L}^\dagger\mathcal{L}$ operator is not useful in this case due to the finite object support. However, if the detector and the collimator are aligned at equally spaced angles, then this system has a discrete rotational symmetry. This property can be used to obtain the singular vectors, and thus determine the pseudoinverse operator of this system. The basic idea behind the approach is to evaluate the singular vectors of the system for one particular detector orientation, and then use this rotational symmetry to determine the singular vectors of the complete system. For this system, rotating the detector and collimator by θ_j is equivalent to rotating the object by $-\theta_j$. Let $\mathcal{T}_j$ be a functional transform corresponding to the geometric rotation $\mathcal{R}_j$. Then

$$\mathcal{T}_j t(\boldsymbol{r}) = t(\mathcal{R}_j^{-1}\boldsymbol{r}), \tag{27}$$

for an arbitrary function $t(\boldsymbol{r})$. Let us now denote the LM operator at detector orientation of $\theta = 0$, by $\mathcal{L}_0$, and at $\theta = \theta_j$ by $\mathcal{L}_j$. Also, let us denote the complete system matrix, which includes the LM operators at all the angles, by $\mathcal{L}$. Then, we can show that [5]

$$\mathcal{L}_j = \mathcal{L}_0\mathcal{T}_j^\dagger. \tag{28}$$

Therefore, the adjoint of the $\mathcal{L}_j$ operator is given by

$$\mathcal{L}_j^\dagger = \mathcal{T}_j\mathcal{L}_0^\dagger. \tag{29}$$

The adjoint is a back-projection operation that smears the 1-D projection data acquired by the detector back into the 2-D space described by the object support. Therefore, performing the backprojection operation for the data acquired at all the angles amounts to summing up all the backprojections. This leads to an easy representation for the backprojection operator:

$$\mathcal{L}^\dagger = \sum_j \mathcal{L}_j^\dagger = \sum_j \mathcal{T}_j\mathcal{L}_0^\dagger. \tag{30}$$

The expression for the $\mathcal{L}^\dagger\mathcal{L}$ operator is then given by [5]

$$\mathcal{L}^\dagger\mathcal{L} = \sum_j \mathcal{L}_j^\dagger\mathcal{L}_j = \sum_j \mathcal{T}_j\mathcal{L}_0^\dagger\mathcal{L}_0\mathcal{T}_j^\dagger. \tag{31}$$

Therefore, we observe that the $\mathcal{L}^\dagger\mathcal{L}$ operator can be expressed in terms of the $\mathcal{L}_0$ operator. The $\mathcal{L}_0$ operator can also be thought of as the planar imaging system operator. Currently, we are investigating that given this relation, how the singular vectors of $\mathcal{L}_0$ and $\mathcal{L}$ operator are related. We can show that when $\hat{p} = p$, i.e. we estimate the true value of the position of interaction, the singular vectors, and thus the pseudoinverse of the $\mathcal{L}$ operator can be found by following a similar approach as in Davison et al. [8]. However, we need to perform further investigation to derive the reconstruction approach when we account for estimation statistics.

V. Conclusions

In this paper, we have investigated the problem of reconstructing object functions from LM data. We have first suggested a general framework to perform this reconstruction, and then applied this framework to an LSIV imaging system with Gaussian PSF, and to SPECT imaging systems. We have presented the reconstruction solution for a SPECT system with infinite object support and infinite angular sampling. We have also shown that for finite angular sampling but infinite support, the reconstruction cannot be performed using the proposed framework. Finally, we have considered a SPECT system with finite support and finite angular sampling, and shown that for this system, the tomographic LM operator is related to the planar LM operator. We are currently investigating the use of this property to perform the reconstruction for this system. As one of the first investigations on analytic LM reconstruction, we have begun with problems in 2-D tomography, but we are also interested in developing these approaches for 3-D tomography, where this work will be very useful. We are also interested in using the developed reconstruction methods to compare systems that acquire LM data to systems that instead bin the data, by evaluating these systems based on objective measures of image quality. These studies will highlight the usefulness of the information that is not lost when data are stored in LM format.

References

[1] A. K. Jha, E. Clarkson, M. A. Kupinski, and H. H. Barrett, "Joint reconstruction of activity and attenuation map using LM SPECT emission data," in *Proc. SPIE Medical Imaging*, vol. 8668, 2013, pp. 86 681W1–9.

[2] L. Bouwens, R. Van de Walle, H. Gifford, M. King, I. Lemahieu, and R. Dierckx, "LMIRA: list-mode iterative reconstruction algorithm for SPECT," *IEEE Trans. Nucl. Sc.*, vol. 48, no. 4, pp. 1364 –1370, Aug 2001.

[3] L. Parra and H. H. Barrett, "List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET," *IEEE Trans. Med. Imaging*, vol. 17, pp. 228–235, Apr 1998.

[4] L. Caucci and H. H. Barrett, "Objective assessment of image quality. V. Photon-counting detectors and list-mode data," *J. Opt. Soc. Am. A*, vol. 29, no. 6, pp. 1003–1016, Jun 2012.

[5] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Wiley, 2004.

[6] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, pp. R541–578, Aug 2006.

[7] P. P. Bruyant, "Analytic and iterative reconstruction algorithms in SPECT," *J. Nucl. Med.*, vol. 43, no. 10, pp. 1343–1358, Oct 2002.

[8] M. Davison and F. Grunbaum, "Tomographic reconstruction with arbitrary directions," *Comm. Pure Appl. Math.*, vol. 34, no. 1, pp. 77–119, 1981.

Image Artifact Influence on Motion Compensated Tomographic Reconstruction in Cardiac C-arm CT

Kerstin Müller, Chris Schwemmer, Günter Lauritsch, Christopher Rohkohl, Andreas Maier, Hein Heidbüchel, Stijn De Buck, Dieter Nuyens, Yiannis Kyriakou, Christoph Köhler, Rebecca Fahrig, and Joachim Hornegger

***Abstract*—In C-arm CT, electrocardiogram (ECG)-gating of data from a single C-arm rotation provides only a few projections per heart phase for image reconstruction. This view starvation leads to prominent streak artifacts and a poor signal to noise ratio. Motion compensation techniques allow for the use of all acquired data for image reconstruction. Cardiac motion can be estimated by deformable 3-D/3-D registration processed on initial 3-D images of different heart phases. The initial 3-D images are computed from the few, ECG-gated data. In this paper, the sensitivity of the 3-D/3-D registration step to the image quality of the initial images is studied. Different reconstruction algorithms are evaluated for a recently proposed cardiac C-arm CT acquisition protocol. An iterative few-view reconstruction, and a filtered backprojection method (FDK) with and without a bilateral filter are investigated with respect to the final motion compensated reconstruction quality. The algorithms were tested on a phantom and on a porcine model using qualitative and quantitative measures. The phantom projection data and geometry is publicly available and can be downloaded from conrad.stanford.edu/data/heart. The results show minor differences between the three motion compensated reconstructions. For two heart phases a relative root mean square error (rRMSE) of $\approx$ 0.09 and 0.06 and an universal image quality index (UQI) of $\approx$ 0.98 and 0.99 was achieved. The motion compensated reconstructions that use all of the projection images show a clear improvement compared to the initial reconstructions. Given the relatively small differences in final image quality, the algorithm of choice is likely to be the one with smallest computational complexity.**

I. Introduction

A. Purpose of this Work

Today, an angiographic C-arm CT system is standard in interventional cardiology laboratories. By acquiring a set of 2-D high-resolution X-ray images from various directions a 3-D image can be computed. Due to the long acquisition times of several seconds, 3-D imaging of moving objects such as the heart is still an open problem. Commonly, an electrocardiogram (ECG) signal is recorded synchronous with the acquisition and a relative heart phase can be assigned to each projection. In order to improve temporal resolution, the reconstruction can be performed with the subset of the projections that lie inside a certain ECG window centered at the favored heart phase [1]. However, the available number of projections is insufficient for imaging of the heart chambers. Streak artifacts hamper the use of the reconstructed volumes. One possible solution is the use of all acquired projection data in combination with compensation for the cardiac motion in the reconstruction step. The cardiac motion can be estimated by registration of initial 3-D volumes of each heart phase to one reference heart phase. The goal of this paper is to find a suitable reconstruction algorithm for the initial 3-D volumes that can provide image quality sufficient for 3-D/3-D registration.

K. Müller, C. Schwemmer, A. Maier and J. Hornegger are with the Pattern Recognition Lab, Department of Computer Science and the Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. email: kerstin.mueller@cs.fau.de. C. Rohkohl, G. Lauritsch, Y. Kyriakou and C. Köhler are with the Siemens AG, Healthcare Sector, Forchheim, Germany. H. Heidbüchel, S. De Buck and D. Nuyens are with the Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium. R. Fahrig is with the Department of Radiology, Stanford University, Stanford, CA, USA. The authors gratefully acknowledge funding support from the NIH grant R01 HL087917 and of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

B. State-of-the-Art

Motion estimation is already investigated in the area of CT imaging. Cardiac motion is calculated by using 3-D/3-D registration of initial images. The deformation of the heart between heart phases is computed by various optimization algorithms. The individual algorithms differ in the objective function, constraints and optimization techniques [2],[3]. In C-arm CT the reconstruction of initial images at different heart phases with projection data acquired during one single C-arm sweep is still an unsolved problem. The reconstruction quality of the initial images is highly dependent on the choice of the acquisition protocol. In recent studies, an image acquisition of multiple-sweeps of the C-arm is used [1], [4]. The number of gated projection images increases and few-view artifacts are avoided. Techniques of 3-D/3-D registration can be applied to estimate the cardiac motion. However, the longer imaging time results in a higher contrast burden and radiation dose for the patient. Therefore, a new protocol for cardiac C-arm CT was presented [5]. It is a single sweep protocol with 10 - 15 s scan time. The quality of the reconstructed images is still critical, even when using compressed sensing algorithms [6]. In this paper, we investigate whether initial images might be generated from this protocol that are of sufficient quality for cardiac motion estimation.

II. Methods and Materials

A. Initial 3-D Image Generation

1) ECG Selection: The ECG-gating is performed by inserting a weighting function with respect to the relative heart phase into the standard FDK approach. The weighting function is centered at a specific heart phase and has the shape of a cosine or rectangular window [7]. Here, we use a strict rectangular gating function of minimal width, i.e. only one view per heart cycle is considered. A certain number H of volumes $f_h(\boldsymbol{x})$, with $h = 1, \ldots, H$ at specific heart phases are reconstructed. Every heart phase h corresponds to a relative heart phase of $[0\%, \ldots, 100\%]$ between two successive R-peaks[1].

2) 3-D Image Reconstruction:

a) ECG-gated Filtered Backprojection Volume Reconstruction (FDK-VR): For this approach, the volumes are reconstructed with a FDK reconstruction algorithm. The ECG-gated FDK images are highly corrupted by noise and have severe streak artifacts.

b) Filtered ECG-gated Filtered Backprojection Volume Reconstruction (FFDK-VR): The FDK-VR volumes in paragraph II-A2a are filtered by a 3-D bilateral filter [8] to reduce the streak artifacts and eliminate noise. The edge-preserving bilateral filter can be applied due to the high contrast inside the heart chambers compared to the streak artifacts.

c) Few-view Volume Reconstruction (F-VR): Images are reconstructed with an iterative few-view reconstruction algorithm that considers the sparse sampling condition. Here, the prior image constrained compressed sensing (PICCS) and the improved total variation (iTV) algorithm are used [9], [10]. The optimization is performed iteratively with a gradient descent scheme and with the same parameters as described in [6]. The resulting volumes have fewer streak artifacts, but are smoother than a standard non-gated FDK reconstruction.

B. Motion Field Estimation via 3-D/3-D Registration

For cardiac motion estimation, one heart phase needs to be selected as reference phase. The corresponding volume is called reference volume and all other volumes are registered to the reference volume. In this paper, a toolbox for nonrigid registration of medical images called `elastix` is used for the 3-D/3-D motion estimation [11]. Here, the deformable registration is based on a uniform cubic B-spline. A multi-resolution scheme of 4 levels is used with a sampling factor of 2 on each pyramid level. A number of $c = 16$ control points in each dimension are used at the highest image resolution. The negative normalized cross correlation (NCC) is used as the objective function and is minimized with an adaptive stochastic gradient descent optimizer. Empirical experiments showed that 500 iterations on each pyramid level are sufficient to result in a minimal objective function value. In order to restrict the motion vector field to a certain region of interest (ROI) where the heart motion is expected, a motion mask delimits the motion. In this first implementation the mask volume is generated manually by the user.

C. Motion Compensated Reconstruction

For final image reconstruction, motion is compensated using Schäfer's method [12]. The resulting volumes are denoted by the type of ECG-gated volume reconstruction with the subscript r (FDK-VR$_r$, FFDK-VR$_r$, F-VR$_r$,).

III. Experiments

A. Phantom Model

The presented 3-D/3-D registration approach has been applied to a ventricle data set comparable to the XCAT phantom [13], [14]. It is assumed that all materials have the same absorption behavior as water. The bloodpool density of the left ventricle was set to 2.5 g/cm^3, the density of the myocardial wall to 1.5 g/cm^3 and the blood in the aorta to 2.0 g/cm^3. We simulated data using a clinical protocol with the same parameters as for the porcine model presented in the following Section III-B. Poisson distributed noise was added to the simulated projections such that the noise characteristic of the reconstructed image fits to that of the clinical data. The phantom projection data and geometry is publicly available and can be downloaded from conrad.stanford.edu/data/heart.

B. Porcine Model

The methods were also applied to an experimental data set of a porcine model. Image acquisition was performed using an Artis zee system (Siemens AG, Healthcare Sector, Forchheim, Germany). The acquisition time was 14.5 s capturing 381 projection images with 30 f/s, and an angular increment of 0.52° during one C-arm sweep [5]. The isotropic pixel resolution was 0.31 mm/pixel (0.19 mm in isocenter) and the detector size 1240 $\times$ 960 pixel. The heart rate was synchronized with the framerate of the imaging acquisition through external heart pacing to 131 bpm. A total of 32 images per heart cycle are acquired resulting in a number of reconstructed heart phases $H = 12$. A volume of ~ 150 ml contrast fluid was administered intravenously at a speed of 10 ml/s beginning 5 s before the X-ray rotation was started. Image reconstruction was performed on an image volume of (25.6 cm)^3 distributed on a 256^3 voxel grid.

IV. Results and Discussion

A. Complexity Analysis

The three approaches (Section II-A2) have different computational complexity. The most complex part for the FDK-VR is the backprojection step with a complexity of $\mathcal{O}(N \cdot n^3)$, with n the side length of the volume and N the number of projections. The backprojection is performed on the GPU. The FFDK-VR utilizes the FDK-VR and additionally performs a filtering step. The used bilateral filter is implemented in a straightforward manner on the GPU and has a complexity of $\mathcal{O}(n^3 \cdot r^3)$, where r denotes the filter size ($r = 5$ in this paper). Most parts of the PICCS+iTV algorithm are implemented on the GPU, but the runtime of the iterative F-VR reconstruction algorithm still exceeds the FDK-VR and FFDK-VR because it consists of several forward/backprojection steps and a whole optimization routine.

B. Quantitative Results: Phantom Data

For the dynamic phantom data the 3-D error and a quantitative 3-D image metric can be evaluated. In order to measure only the artifacts introduced by the heart motion, the non-gated FDK reconstruction using all projections of the static heart phantom of the same heart phase is used as gold standard. The error as well as the image quality metric were evaluated inside a mask around the ventricle. The relative root mean square (rRMSE) was used to quantify the 3-D reconstruction error [15]. As a 3-D image quality metric the universal image quality index (UQI) was computed [15]. The UQI ranges from -1 to 1, with 1 as the best overlap between both reconstructions. The results at two different relative heart phases (30%, 80%) are given in Table I. All three motion compensated reconstructions achieve comparable results, and the image quality improved with respect to the initial images.

Table I: The rRMSE and the UQI of the dynamic phantom model for heart phases 30% and 80%. The best values are marked in bold.

30%	rRMSE	UQI	80 %	rRMSE	UQI
FDK-VR$_r$	**0.09**	**0.98**	FDK-VR$_r$	**0.06**	**0.99**
FFDK-VR$_r$	**0.09**	**0.98**	FFDK-VR$_r$	**0.06**	**0.99**
F-VR$_r$	**0.08**	**0.98**	F-VR$_r$	**0.06**	**0.99**
FDK-VR	0.15	0.95	FDK-VR	0.12	0.96
FFDK-VR	0.12	0.97	FFDK-VR	0.08	0.98
F-VR	0.11	0.97	F-VR	0.08	0.98
Non-gated FDK	0.15	0.96	Non-gated FDK	0.08	0.97

C. Visual Inspection

The results of the phantom data are presented in Figure 1. The ground truth at a heart phase of 80% is illustrated in Figure 1a. The non-gated FDK reconstruction has motion blur around the left ventricle and the myocardial wall is hardly visible (Fig. 1b). In Figure 1c, the FDK-VR depicts the myocardial wall, but is severely degraded by noise and streak artifacts. The FFDK-VR and F-VR have less streak artifacts and a lower noise level, but have a smoother image impression (Fig. 1e and 1g). All three motion compensated reconstructions show comparable and good delineation of the left ventricle (Fig. 1d, 1f and 1h). The results of the porcine data in Figure 2 illustrate that the non-gated FDK reconstruction averages over all heart phases, as highlighted by the doubled catheter and blurred endocardium edges (Fig.2a). The FDK-VR displays the sharp contours of the endocardium, however prominent streak artifacts are apparent (Fig.2c). A better result is provided by the FFDK-VR and F-VR reconstruction (Fig. 2e, 2g). However, both exhibit blurred streak artifacts and are severely smoothed. The motion compensated reconstructions yield the best results (Fig. 2d, 2f and 2h).

D. Edge Response Profiles

The edge response functions of the different volumes are illustrated in Figure 3. The edge response profile is computed as mean edge profile of the lines indicated in Figure 2. It can be seen that the non-gated FDK reconstruction blurs the edge between the endocardium and the epicardium. The three registration approaches achieve a reasonably good edge profile.

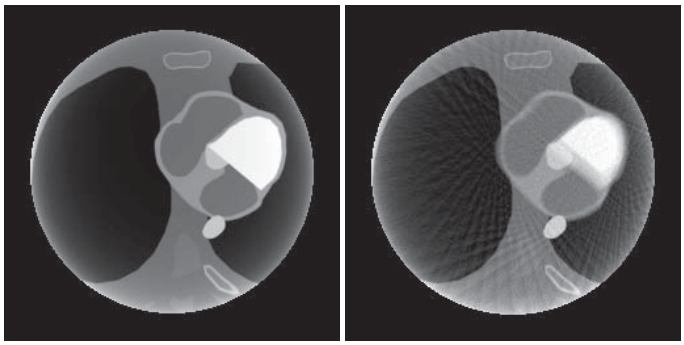
V. Conclusion

We have presented cardiac motion estimation from initial 3-D volume data sets with a deformable B-spline registration. Using motion compensation, the reconstructed image quality is improved compared to the initial image reconstructions. Despite the noise and streak artifacts of the initial images, estimation of a useful motion field is possible. For this image acquisition, non of the presented approaches to enhance the image quality is necessary. This enormously reduces the computational complexity of the framework for dynamic cardiac reconstructions with a C-arm CT system.

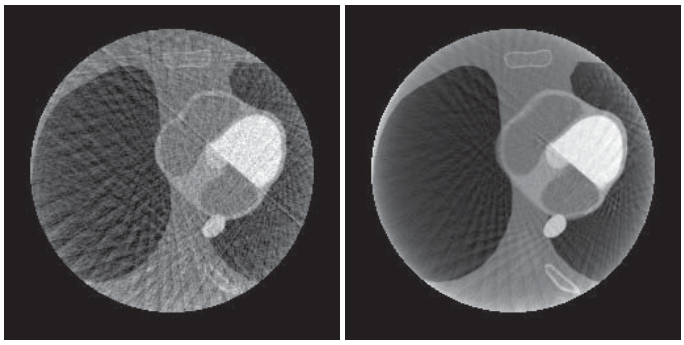
Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

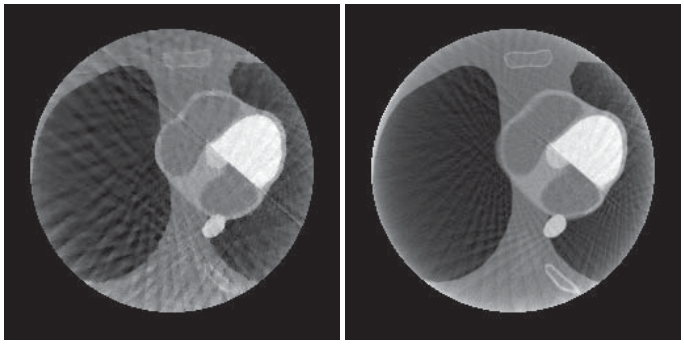
[1] G. Lauritsch, J. Boese, L. Wigström, H. Kemeth, and R. Fahrig, "Towards cardiac C-arm computed tomography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 922–934, July 2006.



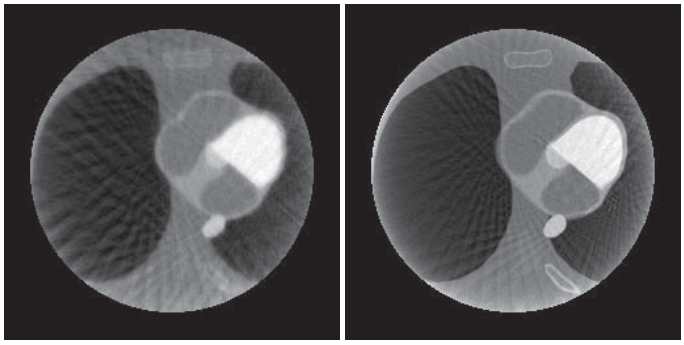
(a) FDK result of static phantom. (b) FDK result of dynamic phantom.



(c) FDK-VR result (32 views). (d) FFDK-VR$_r$ result.

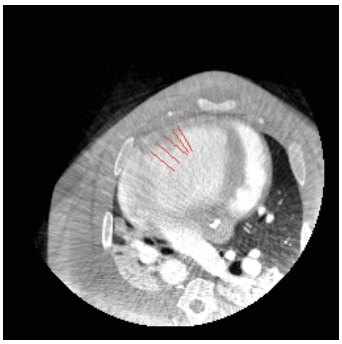


(e) FFDK-VR result. (f) FFDK-VR$_r$ result.

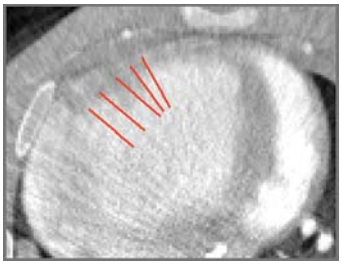


(g) F-VR result. (h) F-VR$_r$ result.

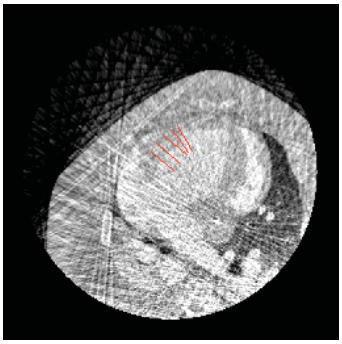
Figure 1: Central slice of initial volumes and motion compensated reconstructions of a phantom model and a relative heart phase of $\approx$ 80%.



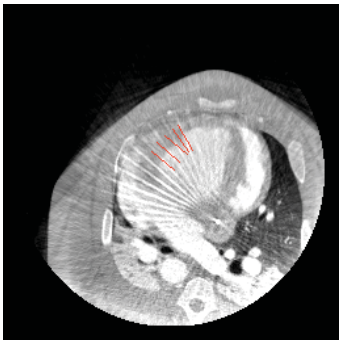
(a) FDK reconstruction.



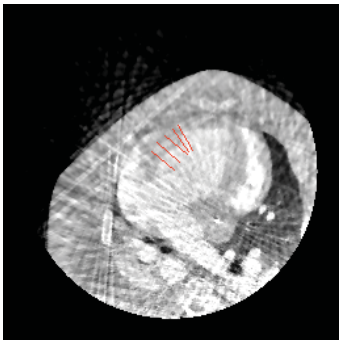
(b) Zoomed line profiles.



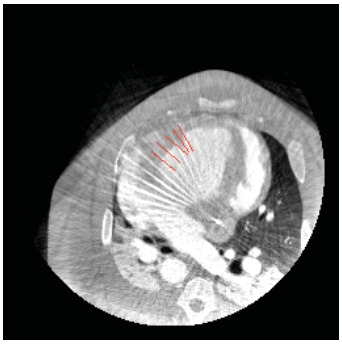
(c) FDK-VR result (32 views).



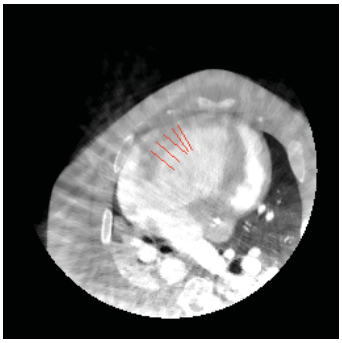
(d) FDK-VR$_r$ result.



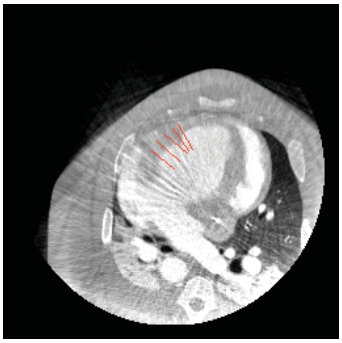
(e) FFDK-VR result.



(f) FFDK-VR$_r$ result.



(g) F-VR result.



(h) F-VR$_r$ result.

Figure 2: Central slice of initial volumes and motion compensated reconstructions of a porcine model and a relative heart phase of $\approx$ 30% (W 1630 HU, C 50 HU, slice thickness 1 mm).

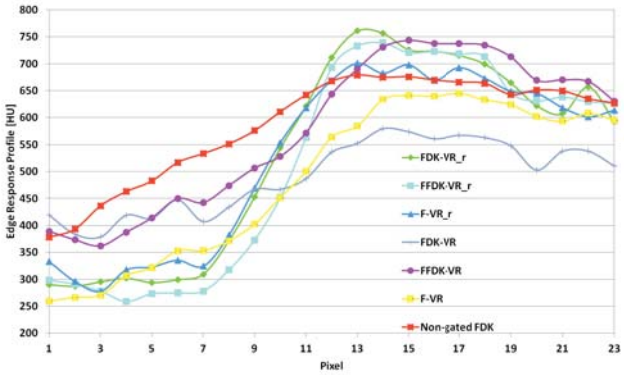


Figure 3: Averaged edge response profile for porcine data of the different algorithms and at 30% heart phase.

[2] A. Isola, M. Grass, and W. Niessen, "Fully automatic nonrigid registration-based local motion estimation for motion-corrected iterative cardiac CT reconstruction," *Med. Phys.*, vol. 37, no. 3, pp. 1093–1109, March 2010.

[3] J. Cammin, P. Khurd, A. Kamen, Q. Tang, K. Kirchberg, C. Chefd Hotel, H. Bruder, and K. Taguchi, "Combined motion estimation and motion compensated fbp for cardiac CT," in *11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2011, pp. 136–139.

[4] M. Prümmer, J. Hornegger, G. Lauritsch, L. Wigström, E. Girard-Hughes, and R. Fahrig, "Cardiac C-Arm CT: A unified framework for motion estimation and dynamic CT," *IEEE Trans. Med. Imaging*, vol. 28, no. 11, pp. 1836–1849, November 2009.

[5] S. De Buck, D. Dauwe, J.-Y. Wielandts, P. Claus, C. Koehler, Y. Kyriakous, S. Janssens, H. Heidbuchel, and D. Nuyens, "A new approach for prospectively gated cardiac rotational angiography," in *Proc. of SPIE Med. Imag. 2013*, 2013, p. to be published.

[6] K. Müller, C. Rohkohl, G. Lauritsch, C. Schwemmer, H. Heidbüchel, S. De Buck, D. Nuyens, Y. Kyriakou, C. Köhler, and J. Hornegger, "4-D motion field estimation by combined multiple heart phase registration (CMHPR) for cardiac C-arm data," in *Proc. of the IEEE NSS / MIC, 2012*, 2012.

[7] C. Rohkohl, G. Lauritsch, A. Nöttling, M. Prümmer, and J. Hornegger, "C-arm ct: Reconstruction of dynamic high contrast objects applied to the coronary sinus," in *Proc. of the IEEE NSS / MIC, 2008*, Dresden, October 2008.

[8] C. Tomasi, "Bilateral filtering for gray and color images," in *Proc. of the IEEE Int. Conf. on Comp. Vis.*, 1998, pp. 839–846.

[9] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, February 2008.

[10] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, pp. 1545–1562, Februar 2011.

[11] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim, "elastix: a toolbox for intensity based medical image registration," *IEEE Trans. Med. Imaging*, vol. 29, no. 2010, pp. 196–205, 2010.

[12] D. Schäfer, J. Borgert, V. Rasche, and M. Grass, "Motion-compensated and gated cone beam filtered back-projection for 3-D rotational x-ray angiography," *IEEE Trans. Med. Imaging*, vol. 25, no. 7, pp. 898–906, July 2006.

[13] W. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, "Realistic ct simulation using the 4D xcat phantom," *Med. Phys.*, vol. 35, no. 8, pp. 3800–3808, August 2008.

[14] A. Maier, H. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast simulation of X-ray projections of spline-based surfaces using an append buffer," *Phys. Med. Biol.*, vol. 57, no. 19, pp. 6193–6210, October 2012.

[15] P. Theriault-Lauzier, J. Tang, and G.-H. Chen, "Prior image constrained compressed sensing: Implementation and performance evaluation," *Med. Phys.*, vol. 39, no. 1, pp. 66–80, January 2012.

4-D Reconstruction of SPECT ^{123}I-MIBG Data Acquired with Slow-Rotation Cameras via Reduction in Spatial and Temporal Dimensions

Yunlong Zan, Rostyslav Boutchko, Qiu Huang, and Grant T. Gullberg

***Abstract*—Dynamic ^{123}I-MIBG SPECT imaging is capable of assessing the severity of heart disease and prognosis in patients with heart failure. However, it is not widely applied in clinics for several reasons, for instance, the lack of reconstruction algorithms with both high accuracy and high efficiency. The goal of this work is to develop a fast algorithm to accurately quantify the metabolism of ^{123}I-MIBG in a clinical SPECT system with slow-rotation cameras, where the greatest challenge of the quantitative analysis involves the inconsistent underdetermined projection data and the high computational load. The novel algorithm presented in this work obtains the blood input function and myocardium time activity curve directly from projections by utilizing the spatial information from the static reconstruction and the nonuniform temporal B-spline basis functions to reduce the spatial and temporal dimensions. Then the kinetic parameters are estimated through a compartmental model with the extracted time activity curves. Compared with the conventional dynamic SPECT reconstruction method without spatial dimension reduction in both phantom simulations and rat experiments, the proposed method provides less-biased time activity curves and more accurate kinetic parameters with less computation time, which makes it practical for small animal studies using clinical systems with slow camera rotation.**

***Index Terms*—Dynamic SPECT, slow-rotation, dimension reduction, B-spline, ML-EM.**

I. Introduction

^{123}I-MIBG is an analogue of the norepinephrine uptake in the presynaptic portion of the sympathetic neurons that innervate the heart [1]. Dynamic ^{123}I-MIBG SPECT reflects sympathetic nervous integrity and predicts clinical outcome in patients with chronic heart failure [2]. Conventionally, the dynamic study includes two steps. Firstly, the time-activity curves (TACs) of the tracer in different tissues are extracted from region of interests (ROIs) on a series of reconstructed images. Secondly, the metabolic parameters are estimated through the kinetic modeling with the TAC of radioactive tracer in the arterial plasma as an input function. The input function is acquired from the frequent blood samplings after the injection of the radioactive tracer. This tedious invasive procedure is not trivial, especially for small animal studies. Alternatively, many researchers have derived the input function from the time series of reconstructed images, which is valid for PET and SPECT with stationary detectors. However, it is not effective in the SPECT system with slow rotating cameras, due to the lower detection efficiency and the slow rotation of the camera. Therefore the extraction of TACs directly from the projection is proposed for a SPECT system with slow rotating cameras [3–7]. Nontheless, large number of parameters to be estimated in these methods not only degrade the accuracy, but also require high computation expenditure.

In this paper, we propose a new algorithm to estimate the TACs using the prior spatial knowledge from the segmentation of the static reconstructed image to reduce the spatial dimension of the dynamic reconstructed image, as well as using the quadratic B-spline to reduce the temporal dimension of the dynamic reconstructed image. The spatial and temporal dimensions reduction increases the constraints on the process of TACs estimation, improves the accuracy of the TACs and the kinetic parameters, and reduces the computation time. This method is validated in a computer simulation and processing data from rat studies of ^{123}I-MIBG, where ^{123}I-MIBG was used to evaluate the change in neural activity in the heart of a Wistar-Kyoto (WKY) normal rat and in a spontaneously hypertensive rat (SHR) with the progression of left ventricular hypertrophy. The difference between the uptake and washout of the tracer in WKY and SHR reflects the severity of cardiac hypertrophy. The proposed method in this paper combined with the SPECT imaging of ^{123}I-MIBG can be translated for patient imaging for improved diagnosis and management of therapy for patients with heart failure.

This paper is organized as follows. Section II demonstrates the algorithm, including the 3-D static reconstruction, spatial dimension reduction, temporal dimension reduction, dynamic reconstruction and compartmental modeling. Section III shows experimental results, where comparisons are made to illustrate the effectiveness of the proposed method. Finally, Section IV concludes the paper.

II. Methods

A. Dual Head Pinhole SPECT system

The dataset for the simulation and the rat study were simulated/acquired using a dual-head pinhole GE Millennium VG3 Hawkeye SPECT/CT scanner with the detector heads arranged in H-mode and equipped with pinhole collimators, as it is shown in Fig 1. For both the simulation study and the rat study, the projection data of 120×88 for each frame

Y. Zan and Q. Huang are with School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China. e-mail:(huangjone@yahoo.com).

R. Boutchko and G. T. Gullberg are with Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

were acquired for 100 min in 1-sec time frames with an angular step of 2 degrees per frame, which makes the projection data set to be 120×88×180×66.

Fig. 1: Clinical dual-head SPECT/CT scanner with custom pinhole collimators

B. 3-D Late Static Reconstruction

To locate the left ventricular blood pool and myocardium tissue, the static image was reconstructed from the acquired late data summed from 1.5 min to 100 min after injection using maximum likelihood expectation maximization (ML-EM) [8] algorithm. The static reconstructed image was then segmented to construct the sparse basis matrix used for 4-D dynamic reconstruction.

C. 4-D Dynamic Reconstruction

1) Spatial dimension reduction

By assuming the radioactivity in each functional region of the image was distributed uniformly, the static SPECT image was segmented to different uniform regions, which allowed us to construct the sparse basis matrix in order to rewrite the conventional SPECT system model $P = FX$ as

$$P = FX = F\Psi S = \Theta S, \tag{1}$$

where Ψ is the $N \times M$ sparse basis matrix constructed from the segmentation of the 3-D static reconstructed image. The column vector in Ψ corresponds to the n-th functional region, and the nonzero elements in this vector represent the membership of these voxels belonging to this functional region. S is the $M \times 1$ column vector of the sparse coefficients. Θ is an $I \times M$ transform coding matrix, which is used in the 4-D reconstruction. Since $M \ll N$, the matrix Θ reduces the number of unknown parameters from N to M, which reduces the problem of underdetermination and thus results in improved reconstruction accuracy and decreases computational complexity.

To reduce the error propagated to the 4-D reconstruction from the inaccurate segmentation due to the poor spatial resolution, we separated the sparse basis matrix into two sub-matrices: $\Theta = [R, O]$. The column vectors in R represented those functional regions segmented with a high confidence, while the column vectors in O represented those ambiguous regions, such as the transitional areas between different functional regions. We used the Fuzzy C-means algorithm [9] combined with a manual modulation to determine the membership value of each voxel to these regions represented by the vectors in R. If the voxel had a high membership value to a certain region in R, then the corresponding element in R was set to be 1. Otherwise, the voxel was considered to belong to a transitional area between different functional regions, which determined a nonzero element in O.

The Fuzzy C-means algorithm [9] is a soft classification method based on the fuzzy theory. In standard classification methods, the voxels are enforced to be inside or outside of a certain functional region with a binary characteristic function,

$$u_{cj} = \begin{cases} 1, & if \quad x_j \in Region(c) \\ 0, & otherwise, \end{cases} \tag{2}$$

while in the soft classification methods, the characteristic function is generalized to be a 'membership function', of which the elements are between zero and one rather than a binary value: $0 \leq u_{cj} \leq 1$. The sum of the membership value u_{cj} for each voxel belonging to all regions is 1,

$$\sum_{c=1}^{C} u_{cj} = 1, j = 1, 2, 3, \ldots, N. \tag{3}$$

The algorithm is an iterative process to find the optimal cluster center and calculate the optimal membership of every voxel that will minimize the objective function. The objective function is,

$$J = \sum_{j=1}^{N} \sum_{c=1}^{C} u_{cj}^{q} d_{cj}^{2}, \tag{4}$$

where $d_{cj}^2 = ||x_j - v_c||^2$ represents the standard Euclidean distance; v_c is the cluster center of the c-th fuzzy division; q is a positive integer (for instance, 2 in this study) to control the fuzzy degree of cluster results. In order to minimize the objective function J, let

$$\begin{cases} \frac{\partial J}{\partial u_{cj}} = 0 \\ \frac{\partial J}{\partial v_c} = 0. \end{cases} \tag{5}$$

Deducing from equation (7), the membership u_{cj} and cluster center v_i are updated in the iterations by,

$$u_{cj} = 1 / \sum_{l=1}^{C} (\frac{d_{cj}}{d_{lj}})^{\frac{2}{q-1}}, \tag{6}$$

$$v_c = \sum_{j=1}^{N} u_{cj}^{q} x_j / \sum_{j=1}^{N} u_{cj}^{q}. \tag{7}$$

2) Temporal dimension reduction

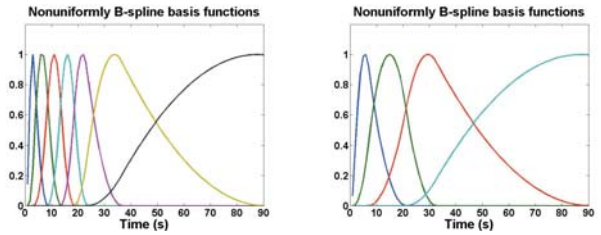
The process of temporal dimension reduction uses the B-splines to reduce the number of the unknown parameters in the temporal dimension. The dynamic system model for SPECT $P(t) = \Theta \times S(t)$ is rewritten after the temporal dimension reductions as

$$P(t) = \Theta \times S(t) = \Theta \sum_{k=1}^{K} C_k B_k(t), \tag{8}$$

where $P(t)$ is the projection data at time t, $S(t)$ is the coefficients of the transform coding matrix at time t, $B(t)$ is

the B-spline basis function. The nonuniform sampling intervals of the B-spline basis functions in this paper were 0-1s, 1s-3s, 3s-8s, 8s-13s, 13s-18s, 18s-23s, 23s-36s and 35-90s for the simulation studies, as well as 0-1s, 1s-6.5s, 6.5s-20.5s, 20.5s-32s, 32s-90s for the rat studies, as shown in Fig. 2. C_k is the coefficient of the k-th B-spline basis function. Thus, reconstruction of $S(t)$ is replaced by reconstructing C_k.

Since $K \ll T$ and $M \ll N$, the spatial and temporal dimensions reduction method reduces the number of the unknown parameters from $N \times T$ to $M \times K$, which improves the accuracy of the reconstructed TACs and reduces the computational complexity.



(a) (b)

Fig. 2: B-spline basis functions for the simulation study (a) and rat studies (b)

3) 4-D reconstruction

The unknown parameters in the vector $\{C_1, \cdots, C_k\}^T$ after the spatial and temporal dimensions reduction were reconstructed through the modified ML-EM algorithm:

$$\hat{C}_{m,k}^{n+1} = \frac{\hat{C}_{m,k}^{n}}{\sum_i \Theta_{im} B_k(t)} \sum_i \Theta_{im} B_k(t) \frac{P_i(t)}{\sum_{m'} \Theta_{im'} B_k(t) \hat{C}_{m',k}^{n}}. \quad (9)$$

The TACs of different functional regions were retrieved through the vector $\{C_1, \cdots, C_k\}^T$ by

$$S(t) = \sum_{k=1}^{K} C_k B_k(t). \quad (10)$$

D. Compartmental Modeling

The Levenberg-Marquardt method [10] was used to fit the time-activity curves acquired from the dynamic reconstruction to a one-tissue compartmental model (Fig. 3) to obtain a quantitative estimate of the metabolic rate of the tissue. The tissue uptake was modeled as

$$C_e(t) = (1-f_v)K_1 \cdot \int_0^t e^{-k_2(t-\tau)} C_p(\tau) d(\tau) + f_v \cdot C_p(t) \quad (11)$$

where K_1 is the uptake rate, k_2 is the washout rate, and f_v is the fraction of vasculature in the tissue incorporating the effect of spillover from the blood pool to the tissue.

III. Experiments and Results

We compared the proposed method and the 4-D reconstruction method with the same temporal processing without the spatial dimension reduction in a simple phantom study and experimental rat studies in this section. The TACs estimated by the method without spatial dimension reduction were averaged over voxels in the ROI from the same segmentation process constructing the sparse basis matrix in the proposed method. The TACs were first reconstructed with two methods, and then the kinetic parameters were estimated using the same compartmental model.

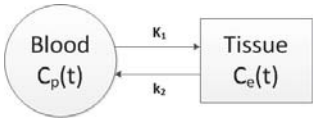


Fig. 3: Two compartmental model

A. Sphere Phantom

The performance of the proposed method and the method without spatial dimension reduction were first compared using a simple 64×64×64-voxel phantom with two spheres (Fig. 4), denoted as A and B. The estimated TACs of region

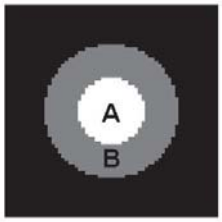
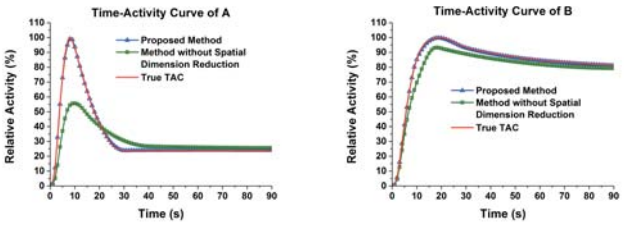


Fig. 4: Sphere phantom

A and region B for the two methods are shown in Fig. 5. The kinetic parameters obtained from the time-activity curves with proposed method and the method without spatial dimension reduction are shown in Table I. It is evident that the proposed method offers more accurate time-activity curves and kinetic parameters, which provides a more reliable estimation of neuronal activity.



(a) (b)

Fig. 5: The true TACs and the estimated TACs of (a) region A and (b) region B by the proposed method and the method without spatial dimension reduction for the sphere phantom

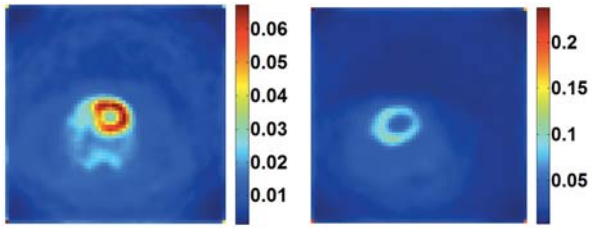
B. Rat Study

The rat data were acquired on a dual-head GE Millennium VG3 Hawkeye SPECT/CT scanner with the detector heads arranged in H-mode and equipped with custom designed pinhole collimators of tungsten with apertures of 1.5×2 mm. In each study, about 5 mCi of ^{123}I-MIBG was

TABLE I: Kinetic parameters of region B in the simulation study estimated from time-activity curves obtained with different dynamic reconstruction methods (1/sec)

Parameter	Ground Truth	Proposed method	Method without Spatial Dimension Reduction
K_1	0.0600	0.0600	0.1012
k_2	0.0300	0.0294	0.0491

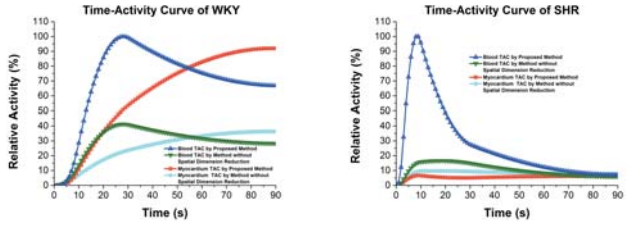
administrated into the rat with the simultaneous dynamic data acquisition. Data were acquired for 100 min in 1-sec time frames with an angular step of 2 degrees per frame. The late data acquired 1.5-100 min after injection were summed and reconstructed. The 3-D static reconstructed images are shown in Fig. 6 for a WKY normal and an SHR.



(a) WKY normal (b) SHR

Fig. 6: The late 3-D static transverse images of rat heart

The estimated TACs of the blood pool and the myocardium with the two methods are shown in Fig. 7. Table II lists the kinetic parameters of the WKY normal and the SHR estimated from the TACs of the two methods, which indicates that the SHR has a slower washout than the WKY normal; however, the rate of uptake is much slower indicating poor chance of survival. As in this case the rate of uptake can be a stronger indicator of survival than the washout rate constant. These results clearly indicates changes in neuronal activity during the development of heart failure. Comparing the kinetic parameters obtained with the two methods, the difference of kinetic parameters between WKY normal and SHR from the proposed method is much more obvious than that from the method without spatial dimension reduction. As a result, this implies that the proposed method can provide a more accurate and stable indicator for heart failure.



(a) (b)

Fig. 7: The estimated TACs of WKY (a) and the SHR (b) by the proposed method and the method without spatial dimension reduction for the rat studies

TABLE II: Kinetic parameters of the myocardium in the rat study estimated from time-activity curves obtained with different dynamic reconstruction methods (1/sec)

Rat	Parameter	Proposed method	Method without Spatial Dimension Reduction
WKY	K_1	0.0262	0.0244
	k_2	0.0120	0.0111
SHR	K_1	0.0026	0.0140
	k_2	0.0013	0.0095

IV. Conclusion

In this paper, we utilized the spatial and temporal dimensions reduction method to improve the accuracy and reduce the computational complexity of the dynamic reconstruction. This method provides more accurate input functions for compartmental modeling, through which the uptake and washout of ^{123}I-MIBG in the myocardium can be quantified accurately, offering a reliable approach to assess the severity of heart disease and prognosis in rats with heart failure.

Acknowledgment

This work was supported in part by the Innovation Program of Shanghai Municipal Education Commission 13ZZ017, China and by NIH grants R01 HL50663 and R01 EB07219 and by the Director, Office of Science, Office of Biological and Environmental Research of the US Department of Energy under contract DE-AC02-05CH11231.

References

[1] G. T. Gullberg, B. W. Reutter, A. Sitek, J. S. Maltz and T. F. Budinger, "Dynamic single photon emission computed tomography-basic principles and cardiac applications", *Phys. Med. Biol.*, vol. 55, No. 20, pp. R111-R191, 2010.

[2] T. Arimoto, Y. Takeishi, A. Fukui, H. Tachibana, N. Nozaki, O. Hirono, H. Yamaguchi, M. Itoh, T. Miyamoto, H. Takahashi, A. Okada, K. Takahashi, I. Kubota, "Dynamic ^{123}I-MIBG SPECT reflects sympathetic nervous integrity and predicts clinical outcome in patients with chronic heart failure", *Ann. Nucl. Med.*, vol. 18, No. 2, pp. 145-150, 2004.

[3] T. Farncombe, A. Celler, D. Noll, J. Maeght, R. Harrop, "Dynamic SPECT imaging using a single camera rotation (dSPECT)", *IEEE Trans. on Nucl. Sci.*, vol. 46, No. 4, pp. 1055-1061, 1999.

[4] T. Farncombe, A. Celler, C. Bever, D. Noll, J. Maeght and R. Harrop, "The incorporation of organ uptake into dynamic SPECT (dSPECT) image reconstruction", *IEEE Trans. on Nucl. Sci.*, vol. 48, No. 1, pp. 3-9 2001.

[5] T. Humphries, A. Celler, M. Trummer, "Slow-rotation dynamic SPECT with a temporal second derivative constraint", *Med. Phys.*, vol. 38, No. 8, pp. 4489-97, 2011.

[6] A. Sitek, G. T. Gullberg, E. V. R. Di Bella, A. Celler, "Reconstruction of dynamic renal tomographic data acquired by slow rotation", *J. Nucl. Med.*, vol. 42, No. 11, pp. 1704-1712, 2001.

[7] B. W. Reutter, R. H. Huesman, K. M. Brennan, R. Boutchko, S. M. Hanrahan, G. T. Gullberg, "Longitudinal evaluation of fatty acid metabolism in normal and spontaneously hypertensive rat hearts with dynamic microSPECT imaging", *Internal. J. Mol. Imag.*, vol. 2011, 2010.

[8] L. A. Shepp, Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", *IEEE Trans. on Med. Imaging*, vol. 1, No. 2, pp. 113-122, 1982.

[9] J. C. Bezdek, R. Ehrlich, W. Full, "FCM: The fuzzy c-means clustering algorithm", *Computers & Geosciences*, vol. 10, No. 2, pp. 191-203, 1984.

[10] K. Levenberg, "A method for the solution of certain problems in least squares", *Quart. of Appl. Math.*, vol. 2, pp. 164-168, 1944.

A Geometric Calibration Method for an Open Cone-Beam CT System

Fabian Stopp, Adam J. Wieckowski, Marc Käseberg, Sebastian Engel, Felix Fehlhaber, Erwin Keeve

***Abstract*—The image quality of cone-beam CT systems depends directly on the precise knowledge of position and orientation of the X-ray source and the detector. The current methods to determine this geometric information are mainly focused on conventional cone-beam CTs with planar or near-planar scanning trajectories. Due to the fixed alignment of X-ray source and detector, such systems have disadvantages in intraoperative use.**

Therefore, we develop a first prototype for cone-beam CT characterized by a free alignment of X-ray source and detector. This results in an open system allowing an intraoperative access to the patient and the implementation of non-planar scanning trajectories in the operating room.

In this paper, we present a geometric calibration method to determine the position and orientation of X-ray source and detector for any arbitrary projection. Enhancing the theoretical method proposed in Mennessier et al. [1] by introducing an asymmetrical marker arrangement, we realized a calibration method suitable for practical use. We analyzed the resulting accuracy and applied our approach to the open cone-beam CT prototype.

***Index Terms*—Computed tomography, cone-beam, geometric calibration, non-planar scanning trajectories.**

I. Introduction

In cone-beam computed tomography (CBCT) an object can be three-dimensionally (3D) reconstructed by acquiring X-ray images of this object from different directions. Conventional CBCT systems are characterized by a rigid configuration of X-ray source and image detector, mostly fixed on a C-shaped arm or inside a closed gantry. For 3D image acquisition, source and detector are rotating on a planar trajectory around the patient. With such systems a high image quality is achievable, but the intraoperative use during surgery can be time consuming and complicated. Due to the fixed arrangement of X-ray source and image detector on opposite sides, the patient is surrounded by the system setup and the access for the surgeon is restricted. Therefore, we develop a first experimental open CBCT system for interventional surgery (ORBIT, fig. 1 and 2). The X-ray source is fixed on a robot-arm and the digital flat-panel detector is mounted on a self-constructed motorized mechanism directly connected to the patient table. This system allows a free alignment of X-ray source and image detector towards the patient and offers new opportunities for non-planar scanning trajectories (e.g. fig. 2).

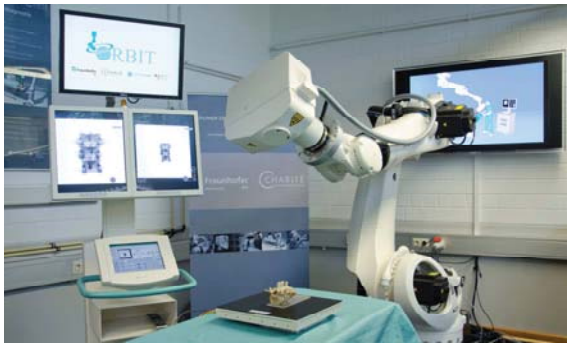


Fig. 1. Current prototype of the open cone-beam CT system (ORBIT) consisting of a pulsed X-ray source (Ziehm Vision R 20 kW) on a robot arm (Kuka KR 150 R2700 extra) and a currently fixed flat-panel detector (Varian PaxScan 3030+).

The essential precondition for CBCT is the knowledge of the exact projection geometry of each acquired image. Therefore the position of the focal spot of the X-ray source and the position and orientation of the X-ray image detector is needed. This information, described by nine parameters, can be determined by a geometric calibration. Most of the available calibration methods use dedicated objects with a known geometric configuration of small balls of high attenuation. By acquiring X-ray images of these objects and identifying the ball projections, the needed parameters of each single image can be determined. But the majority of proposed methods were developed for conventional CBCT systems with planar or near-planar scanning trajectories (e.g. in [2]-[7]). Other approaches without constraints on the scanning trajectory or the alignment of source and detector use complex numerical optimization techniques, e.g. in [8]. In [1] a fully analytical calibration method for near-planar trajectories (using a six points calibration phantom) and for arbitrary scanning trajectories (using a 14 points calibration phantom) were introduced and first simulated results were shown. In further work the direct calibration method for near-planar trajectories was realized and applied to an isocentric c-arm X-ray system using a 6 balls calibration object [9].

Based on the work presented in Mennessier et al. [1], we developed, applied and evaluated a direct geometric

This work is funded by the German Federal Mininstry of Education and Research (BMBF), research grant 13EZ1115A-C.

F. Stopp and Prof. Dr. E. Keeve are with the Department of Maxillofacial Surgery and Clinical Navigation, Charité – Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany (e-mail: keeve@charite.de).

A. J. Wieckowski, M. Käseberg, S. Engel, F. Fehlhaber and Prof. Dr. E. Keeve are with the Fraunhofer-Institute for Production Systems and Design Technology IPK, Pascalstrasse 8-9, 10587 Berlin, Germany.

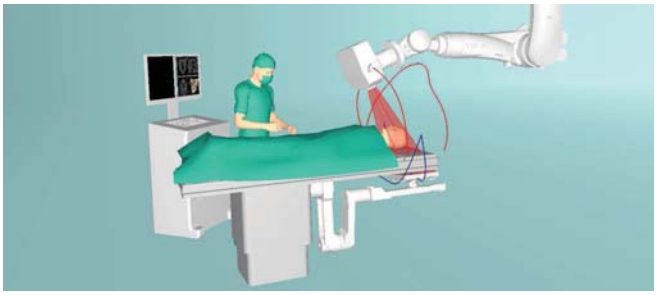


Fig. 2. Open cone-beam CT system ORBIT with an X-ray source fixed on a ceiling-mounted robot arm and a robot-guided flat panel detector directly connected with the patient table. The red path and the blue path indicate an exemplary non-planar scanning trajectory of the X-ray source and detector.

calibration method with a subsequent optimization of the parameters for arbitrary scanning trajectories. In contrast to the simulations in [1], no additional features of the calibration markers are used for marker identification, like absorption coefficients or marker size. After first simulation results, we applied our calibration method to the open CBCT prototype.

II. Our Geometric Calibration Method

To calibrate and determine the nine geometric parameters of an arbitrary image acquisition (X-ray source position, image position and image orientation), we defined the marker arrangement of the calibration object.

A. Marker arrangement

Similar to [1], we align four ball markers on each of the three orthogonal axes and two additional markers on the diagonal axis ($\mathbf{x}_1..\mathbf{x}_4$, $\mathbf{y}_1..\mathbf{y}_4$, $\mathbf{z}_1..\mathbf{z}_4$, $\mathbf{s}_1$, $\mathbf{s}_2$). For the marker assignment to be independent of additional marker features, like size or X-ray absorption coefficients, we defined five constants K_1, K_2, α_x, α_y, and α_z and modified the arrangement of $M = 14$ ball markers (fig. 3). The arrangement of the four markers on the x-axis is:

$$\begin{aligned} \mathbf{x}_1 &= -\alpha_x K_1 \mathbf{e}_x & \mathbf{x}_2 &= -K_1 \mathbf{e}_x \\ \mathbf{x}_3 &= K_2 \mathbf{e}_x & \mathbf{x}_4 &= \alpha_x K_2 \mathbf{e}_x \end{aligned} \tag{1}$$

The arrangement of the eight markers on the y- and z-axis (using α_y and α_z) is equivalent to (1). The three unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$ and $\mathbf{e}_z$ represent the axes of the calibration object's coordinate system. The two diagonal markers are aligned as follows:

$$\mathbf{s}_1 = -K_1(\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z) \quad \mathbf{s}_2 = K_2(\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z) \tag{2}$$

Depending on the system characteristics of our open CBCT setup (image size of 298 x 298 mm² and a cone-beam opening angle of 16°), we used the following values: $K_1 = 25$ mm, $K_2 = 35$ mm and $\alpha_x = 1.4$, $\alpha_y = 1.7$, and $\alpha_z = 2$. These constants fulfill our basic precondition for good calibration results: the 14 marker must be fully contained and fill as much as possible of the X-ray image.

B. Marker detection

To detect the regions of the projected ball markers in the X-ray images, we apply the following four steps:

1. Segmenting the regions of the ball marker projections by an adaptive threshold.

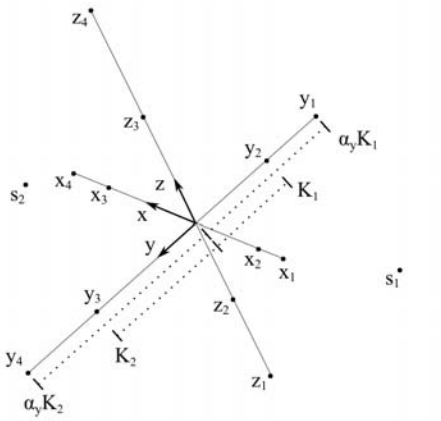


Fig. 3. The arrangement of the 14 ball markers of the calibration object. Four markers are placed on each orthogonal axis and two on the diagonal. The exact positions are defined by the constants K_1, K_2 and α_x, α_y, and α_z.

2. Classification of the segmented regions.
3. Analyzing the blob response of the region using the determinant of Hessian.
4. Analyzing the foreground to background intensity difference.

A region is described by a set of N pixel positions $\mathbf{r}_k$ and intensities I_k. Given the neighborhood of that region with a mean intensity of background b, the projection of the marker center is estimated as the center of mass with background suppression [10]:

$$\mathbf{d} = \left(\sum_{k=1}^{N} (I_k - b) \cdot \mathbf{r}_k \Big/ \sum_{k=1}^{N} (I_k - b) \right) \tag{3}$$

In the following section, we assume that all 14 marker projections are detected in the image. The degenerated cases (e.g. marker overlaps) are described in section II E.

C. Marker assignment

We divide the detected marker projections $\mathbf{d}_i$ ($i = 1..M$) into four groups, representing the three orthogonal axes and the both diagonal marker. From each found line of four markers, we can infer the respective axis of the calibration object. We identify the lines by retrieving an approximation of the constant α from the X-ray image. Assuming $\mathbf{d}_1 \ldots \mathbf{d}_4$ are the four detected marker projections, classified as a line (ordered along that line), and $\mathbf{c}$ is the projection of the calibration object's origin $\mathbf{o}$ (symmedian point of the triangle bound by the three detected lines [1]), an approximation of α_j ($j \in \{x,y,z\}$) is given by:

$$\alpha_j = \frac{1}{2}\left(\frac{\alpha_j K_1}{K_1} + \frac{\alpha_j K_2}{K_2} \right) \approx \frac{1}{2}\left(\frac{\|\mathbf{d}_1 - \mathbf{c}\|}{\|\mathbf{d}_2 - \mathbf{c}\|} + \frac{\|\mathbf{d}_4 - \mathbf{c}\|}{\|\mathbf{d}_3 - \mathbf{c}\|} \right) \tag{4}$$

Using the value α_j obtained from (4), we can determine the corresponding axis of the marker projections. But we cannot infer the order of the markers in terms of direction nor can we assign both diagonal marker projections. Because each group is assigned in two ways (e.g. the projections of the markers on the x-axis: $\{(\mathbf{x}_1, \mathbf{d}_1), (\mathbf{x}_2, \mathbf{d}_2), (\mathbf{x}_3, \mathbf{d}_3), (\mathbf{x}_4, \mathbf{d}_4)\}$ and $\{(\mathbf{x}_1, \mathbf{d}_4), (\mathbf{x}_2, \mathbf{d}_3), (\mathbf{x}_3, \mathbf{d}_2), (\mathbf{x}_4, \mathbf{d}_1)\}$), we are considering 16 possible marker assignments.

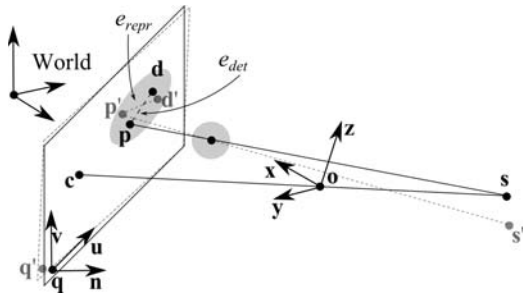


Fig. 4. The projection geometry: X-ray source position **s**, image position **q** and a calibration object at **o** with a ball marker. The X-ray projection of **o** is called **c**. The real image position of the projected marker center is **p**, the detected position is **d** (the difference of both is the detection error e_{det}). Position **p**' is the re-projection of the marker using the calibrated parameters **s**' and **q**' (the difference of **p**' and **d**' is the re-projection error e_{repr}).

D. Geometric parameter determination

We directly determine the geometric parameters for each possible marker assignment by solving a linear equation system as described in [1]. Each set of geometric parameters is verified using a score based on the re-projection errors $e_{repr,i}$ (fig. 4) of this configuration: mean μ_e, variance σ_e^2 and maximum e_{max}. Because of the unambiguity of the marker projections (imposed by the asymmetry constraint $K_1 \neq K_2$), the best score indicates the correct marker assignment and geometric parameter set.

In the last step we refine the geometric parameters using a non-linear Levenberg-Marquardt optimization algorithm to minimize the sum of squared re-projection errors e_{repr}. Because of the good initialization with the directly determined parameters, the optimization problem can be solved efficiently.

E. The degenerated cases

There are two basic problems that might occur when dealing with X-ray images of the marker arrangement described in section II A (for example fig. 5 right):

- Marker overlaps: two or more ball markers form an overlap in the image.
- Structural overlaps: two axes of markers form nearly the same line in the image.

The detection of marker overlaps is not fully possible by using features like size or intensities of detected regions (especially with oblique X-ray projections and elliptical marker regions). We recognize marker overlaps, if less than 14 regions were detected in the image. We then assume for each region that it is an overlap and divide all 14 markers accordingly into four groups (representing the three orthogonal and one diagonal axis). The respective marker centers $\mathbf{d}_i$ of an overlap are approximated by determining the center and the main axis of this region. All marker groups are verified by the following criteria:

- Each detected line must contain exactly the same number of marker projections on each side of **c**.
- We dissolve a structural overlap with eight marker projections, by verifying both identified lines using (4).
- There must be either none or two marker projections not assigned to any line ($\mathbf{s}_1$ and $\mathbf{s}_2$).

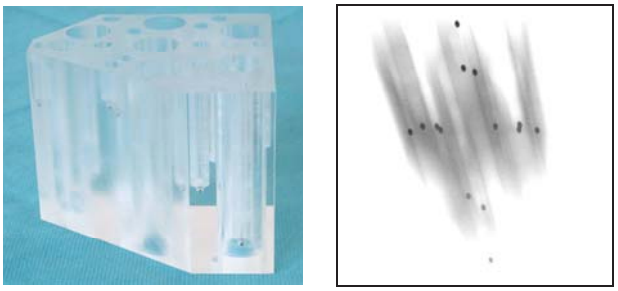


Fig. 5. Calibration object with 14 steel balls (left) and exemplary X-ray image of the calibration object with our CBCT system ORBIT (right).

- In case a marker projection is detected near **c**, we assume that the X-ray projection is in direction of an orthogonal or diagonal axis and these marker projections are ignored.

For each marker group in accordance with the criteria, we perform a marker assignment (section II C) and verify the resulting parameter set using the score function (section II D).

The approximated marker centers $\mathbf{d}_i$ of an overlap are only used for marker assignment and not for the determination of the geometric parameters.

III. Calibration Object

We constructed and manufactured a calibration object with 14 drill-holes to perform a geometric calibration of single X-ray projections images acquired with our open CBCT system. By the defined arrangement and depth of the drill-holes, 14 steel balls with 3 mm diameter can be placed accordingly to the previously described marker arrangement in section II A (Fig. 5 left). To verify and compensate manufacturing inaccuracies, we scanned the calibration object with an industrial CT and measured the exact steel ball locations. By taking arbitrary X-ray projections of this calibration object, the geometric parameters of each image can be determined using our proposed method (fig. 5 right).

IV. Results

At the current project state, the image detector of our open CBCT prototype is fixed on the patient table (fig. 1). Therefore, we first applied our calibration method on a scanning trajectory similar to tomosynthesis: a circularly moving X-ray source above an object on a fixed image detector (fig. 6).

A. Simulation results

Using our simulation environment, we generated 360 artificial projection images of our calibration object equally distributed along this scanning trajectory. The acquisition parameters were defined according to our open CBCT system setup: 298 x 298 mm² image size with 1024 x 1024 px and a distance from X-ray source to image center of approximately 1050 mm. The artificial images are ideal projections of our marker arrangement without noise or motion artifacts. We calculated the nine geometric parameters of all images and compared the results with the defined parameters of the

simulation environment. Additionally, we calculated the accuracy of our marker center detection algorithm (e_{det}) and the re-projection error of the calibrated configuration (e_{repr}). The results are shown in table 1.

TABLE I
SIMULATION RESULTS

		μ	σ	max
Marker detection error e_{det} [mm]		0.009	0.005	0.036
Image center position error [mm]	u	0.029	0.026	0.166
	v	0.033	0.035	0.239
	n	0.046	0.037	0.237
	∥	0.066	0.054	0.344
	⊥	0.004	0.002	0.015
X-ray source position error [mm]	u	0.634	0.548	3.375
	v	0.672	0.722	3.988
	n	0.940	0.770	3.954
	∥	1.375	1.110	5.680
	⊥	0.136	0.104	0.554
Image orientation error [°]	η	0.004	0.004	0.022
	θ	0.005	0.005	0.027
	φ	0.006	0.004	0.040
Re-projection error e_{repr} [mm]		0.008	0.005	0.038

The image center and source position errors are given in image coordinate system **u-v-n**. The symbols ∥ and ⊥ indicate position deviations in direction (∥) and perpendicular (⊥) to the central X-ray beam. The angles η, θ, and φ describe the image rotation errors in **uv**-, **vn**- and **un**-plane of the image coordinate system.

B. Experimental results

Furthermore, we applied the calibration method on the open CBCT prototype and acquired 360 images of our calibration object. We calculated the geometric parameters of each acquired image and repeated the scanning trajectory with vertebral bodies of the lumbar spine. Based on the calculated geometric parameters and the 360 projection images, we reconstructed the scanned volume with a simultaneous algebraic reconstruction technique. Fig. 6 shows the vertebral test bodies and an axial and coronal slice of the reconstructed volume. As a quantitative measurement of the experimental result, we calculated the re-projection errors $e_{repr,i}$ using the calibrated parameters and the found marker assignment of each image of the calibration object: $\mu = 0.058$ mm, $\sigma = 0.047$ mm, $e_{max} = 0.583$ mm.

V. CONCLUSION AND DISCUSSION

The errors in the simulation (shown in table I) result from the marker center detection of the projected ball markers and the parameter determination with our calibration method. Discrepancies between the real and calibrated geometric parameters mainly occur in direction of the projection (indicated with symbol ∥ in table I), but these deviations have little impact on the accuracy of the x-ray projections. Regarding the image plane, the maximum error of the image center position is less than a pixel size of 0.29 mm.

By applying our method on the open CBCT system, additional errors influence the resulting accuracy: the repeatability of the X-ray source positioning by the robot-arm and inaccuracies of the manufactured calibration body. Errors caused by the movement of the X-ray source during the image acquisition are excluded at the moment, as the image acquisition is currently done in defined fixed positions. In further project work we will develop a motorized mechanism to move the flat-panel detector independently of the X-ray source and execute freely definable scanning trajectories. To compensate occurring mechanical instabilities, we will perform an additional online calibration method during the image acquisition.

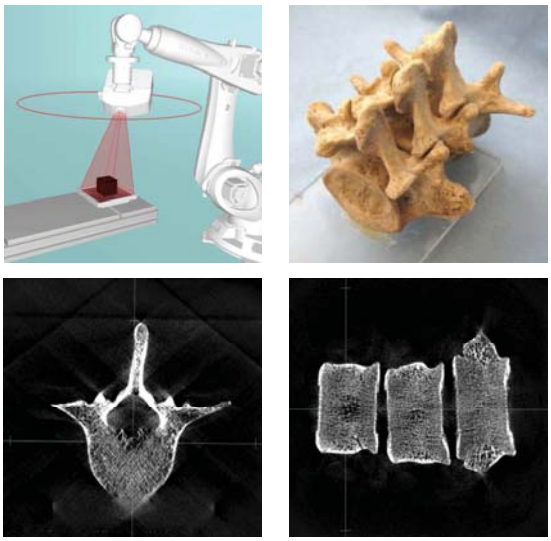


Fig. 6. a) Executed and calibrated scanning trajectory with a fixed flat-panel detector; b) Vertebral bodies of the lumbar spine; c) Axial slice of the reconstructed volume; d) Coronal slice of the reconstructed volume.

REFERENCES

[1] C. Mennessier, R. Clackdoyle, and F. Noo, "Direct determination of geometric alignment parameters for cone-beam scanners.," *Physics in medicine and biology*, vol. 54, no. 6, pp. 1633–60, Mar. 2009.

[2] Y. Cho, D. J. Moseley, J. H. Siewerdsen, and D. a. Jaffray, "Accurate technique for complete geometric calibration of cone-beam computed tomography systems," *Medical Physics*, vol. 32, no. 4, p. 968, 2005.

[3] K. Yang, A. L. C. Kwan, and D. F. Miller, "A geometric calibration method for cone beam CT systems," *Med Phys.*, vol. 33, no. 6, pp. 1695–1706, 2006.

[4] J. C. Ford, D. Zheng, and J. F. Williamson, "Estimation of CT cone-beam geometry using a novel method insensitive to phantom fabrication inaccuracy: Implications for isocenter localization accuracy," *Medical Physics*, vol. 38, no. 6, p. 2829, 2011.

[5] M. J. Daly, J. H. Siewerdsen, Y. B. Cho, D. a. Jaffray, and J. C. Irish, "Geometric calibration of a mobile C-arm for intraoperative cone-beam CT," *Medical Physics*, vol. 35, no. 5, p. 2124, 2008.

[6] X. Wang, J. G. Mainprize, M. P. Kempston, G. E. Mawdsley, and M. J. Yaffe, "Digital breast tomosynthesis geometry calibration," *Proceedings of SPIE*, vol. 6510, p. 65103B–65103B–11, 2007.

[7] S. Johnston, G. Johnson, and C. Badea, "Geometric calibration for a dual tube/detector micro-CT system," *Medical Physics*, vol. 35, no. 5, pp. 1820–1829, 2008.

[8] N. Robert, K. N. Watt, X. Wang, and J. G. Mainprize, "The geometric calibration of cone-beam systems with arbitrary geometry.," *Physics in medicine and biology*, vol. 54, no. 24, pp. 7239–61, Dec. 2009.

[9] C. Mennessier, B. Spencer, R. Clackdoyle, and T. Xu, "Distortion correction, geometric calibration, and volume reconstruction for an isocentric C-Arm X-Ray system," *(NSS/MIC), 2011*, pp. 2943–2947, 2011.

[10] R. Clackdoyle and C. Mennessier, "Centers and centroids of the cone-beam projection of a ball.," *Physics in medicine and biology*, vol. 56, no. 23, pp. 7371–91, Dec. 2011.

Total-variation stokes strategy for sparse-view CT image reconstruction from clinical data

Yan Liu, Hongbing Lu, Hao Zhang, Ke Wang and Zhengrong Liang

***Abstract* – Previous works have shown that computed tomography (CT) images can be reconstructed from sparse-view data by minimizing the constrained TV of the to-be-estimated image. Considering the incompressible velocity field of the image voxels along the tangent directions of the isophote lines, this paper proposes a new total variation stokes (TVS) strategy for CT image reconstruction from sparse-view projection data. In this newly-proposed algorithm, a tangent field is first consolidated for tangent vector estimation, and then a minimization problem based on estimated vector is addressed and resolved in computation. The to-be-estimated image is iteratively solved by this two-step framework with constraints from the data fidelity. By introducing this tangent vector estimation, the effects of staircase and patchy artifacts in the uniform region, which are often observed in TV minimization problem, can be efficiently suppressed without sacrificing edge information. In this study, the TVS method was evaluated by patients' thorax raw data acquired from a clinical multi-slice CT scanner. From the results we observed that the proposed TVS strategy can accurately reconstruct the images and mitigate the patchy artifacts from sparse-views data, comparing to the TV-projection onto convex sets (TV-POCS) method and its general case: adaptive weighted TV-POCS (AwTV-POCS) method. In addition, an improvement was also observed in universal quality index (UQI) measurement study by TVS method compared to AwTV/TV-POCS method. Further evaluations on the proposed TVS using more data are under progress.**

***Index Terms* – total variation, stokes, sparse-view, image reconstruction.**

I. Introduction

Total variation (TV) model has been proved that it could be successfully applied to computed tomography (CT) image reconstruction for solving the sparse-view problems over the last several years [1] [2] [3]. However, the resulting images from the TV minimization are always reported to suffer the effects of undesired staircase or patchy artifacts [3] [4]. Those small patchy artifacts can mimic small lesions which are potentially harmful for medical diagnosis. A recently proposed adaptive weighted total variation-projection onto convex sets (AwTV-POCS) method is an attempt to solve the over-smoothing and patchy artifacts by introducing anisotropic weights to the conventional TV model [5]. The results from AwTV-POCS method show better noise suppression and edge preserving than using the conventional total variation-projection onto convex sets (TV-POCS) method. However, small patchy artifacts are still observed under the extra sparse-view cases. Recently, a total variation-stokes (TVS) method was introduced for solving image denoising and inpainting problems in [6] [7]. The authors demonstrated that the TVS method can efficiently suppress noise and mitigate the patchy artifacts without sacrificing the fine structure of the images. However, due to the lack of prospective images, the TVS method can't be directly utilized for CT image reconstruction. In this study, a modified TVS method is proposed to show its potentials for image reconstruction from sparse-view data. To show our respect to original authors of the TVS method [6] [7], we use the same TVS name. However, it should be noticed that the TVS method introduced in this study has been modified for solving specific reconstruction problem of CT imaging.

In our TVS method, a two-step approach is introduced involving a smoothing along the tangential filed and a surface fitting with constraints in image domain. The motivation of this method is the introduction of a tangential field corresponding to an incompressible velocity field. In addition, in our study, we assume that the tangential field is divergence free, which indicates the possibility to recover missing and noise data along the tangential field. Therefore, the effects of the staircase and patchy artifacts caused by the over-smoothing along the normal direction can be efficiently mitigated.

This paper is organized as follows. In section two, we present the methodology of our new TVS strategy for image reconstruction from sparse-view data. In section three, preliminary results from the proposed TVS methods are shown and compared to the results from conventional AwTV/TV-POCS methods. Finally, a discussion and preliminary conclusion are given in section four.

II. Methods

For a given 2D image $f(\mu)$, where μ is the desired attenuation coefficient in CT image, two orthogonal vectors in image domain: the normal vector $\mathbf{n}$ and the tangential vector $\boldsymbol{\tau}$ of the image are mathematically defined as:

This work was supported in part by the NIH/NCI under Grant #CA143111 and #CA082402. Hongbing Lu is partially supported by the National Natural Science Foundation of China under grant No. 81230035 and by the National Key Technologies R&D Program of China under grant No. 2011BAI12B03. *Asterisk indicates corresponding author.*

Y. Liu is with the Departments of Radiology and Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY, 11794, USA.

H. Lu is with the Department of Biomedical Engineering, Fourth Military Medical University, Xi'An, Shanxi, 710032, China.

H. Zhang is with the Departments of Radiology and Biomedical Engineering, Stony Brook University, Stony Brook, NY, 11794, USA.

K. Wang is with the Department of Applied Mathematics & Statistics, Stony Brook, NY, 11794, USA.

* Z. Liang is with the Departments of Radiology, Computer Science and Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794, USA (telephone: 631-444-7837, e-mail: jerome@mil.sunysb.edu).

$$\mathbf{n} = \nabla f(\mu) = \left(\frac{\partial f}{\partial \mu_x}, \frac{\partial f}{\partial \mu_y} \right)^T \text{ and } \boldsymbol{\tau} = \nabla^{\perp} f(\mu) = \left(\frac{\partial f}{\partial \mu_y}, -\frac{\partial f}{\partial \mu_x} \right)^T, \quad (1)$$

where ∇ denotes the differential operator, T represents the transpose operator, and $\nabla^{\perp}$ is the orthogonal differential operator. The two vectors should satisfy the irrotationality and incompressibility conditions separately, which can be mathematically expressed as [6]:

$$\nabla \times \mathbf{n} = 0 \quad \text{and} \quad \nabla \cdot \boldsymbol{\tau} = 0. \quad (2)$$

The left equation shows that the cross product of the differential operator and vector $\mathbf{n}$ is equal to zero, which indicates the curl of the normal vector is zero. The right equation shows that a dot product of the differential operator and vector $\boldsymbol{\tau}$ is also equal to zero, which indicates the divergence of tangent vector is zero. Unlike the TV-POCS algorithm introduced in [4], which assumes the image is piecewise constant and estimates the desired image along the normal field, the TVS strategy requires to restore the missing data along the tangential field. Although the piece-wise assumption can help us to propagate information to the missing data along the normal field, it can't guarantee the consistency along the tangent directions. Thus, the results from AwTV/TV-POCS always show undesired patchy artifacts. In our proposed TVS method, the image restoring is executed along tangent direction and the consistency along the tangent direction is well considered. Therefore, the undesired patchy artifacts can be eliminated efficiently due to the consistency along both of the normal direction and tangent direction. In this study, we find that this strategy always show a pleasant image and preserve the edges quite well. Inspired by previous study, we choose the two steps iterative framework as introduced in [6] to solve this problem.

In the first step, the tangent vector is estimated by minimizing the TV norm of the tangential vector with incompressibility constraints, which indicates the tangent field is a passive vector field. This step is called as tangent field smoothing (TFS) step in the following. The minimization problem of this step can be written as:

$$\min_{\boldsymbol{\tau}} \int_{\Omega} |\nabla \tau| d\mu \quad \text{subject to} \quad \nabla \cdot \boldsymbol{\tau} = 0 \quad (3)$$

To solve such partial differential equations (PDE), the Augmented Lagrangian (AL) method is utilized [6]. The updating function of Eq. (3) is:

$$v^{n+1} = v^n + \Delta t_1 \left(D_x^- \left(\frac{D_x^+ v^n}{T_1^n} \right) + D_y^- \left(\frac{D_y^+ v^n}{T_2^n} \right) + D_x^- \left(\lambda^n + Div(\tau^n) \right) \right), \quad (4)$$

$$u^{n+1} = u^n + \Delta t_1 \left(D_x^- \left(\frac{D_x^+ u^n}{T_2^n} \right) + D_y^- \left(\frac{D_y^+ u^n}{T_1^n} \right) + D_y^- \left(\lambda^n + Div(\tau^n) \right) \right), \quad (5)$$

$$\lambda^{n+1} = \lambda^n + \Delta t_1 \left(D_x^+ v^n + D_y^+ u^n \right), \quad (6)$$

where

$$Div(\tau^n) = D_x^+ v^n + D_y^+ u^n, \quad (7)$$

$$T_1 = \sqrt{\left(A_x \left(C_y^1 v^n \right) \right)^2 + \left(D_x^+ v^n \right)^2 + \left(D_y^+ u^n \right)^2 + \left(M_y \left(C_x^1 u^n \right) \right)^2 + \varepsilon}, \quad (8)$$

$$T_2 = \sqrt{\left(A_y \left(C_x^1 v^n \right) \right)^2 + \left(D_y^+ v^n \right)^2 + \left(D_x^+ u^n \right)^2 + \left(M_x \left(C_y^1 u^n \right) \right)^2 + \varepsilon}, \quad (9)$$

where u and v are the gradient along y and x directions, Δt_1 is a relaxing parameter which can control the updating step length in the TFS step, $D_x^{\pm}$ and $D_y^{\pm}$ are defined as forward/backward difference operators along x and y directions, C_x^1 and C_y^1 in Eq. (8) and (9) are defined as the first order centered difference operators along x and y directions, M_x and M_y are the first order neighbors' mean along x and y directions, which can be expressed as:

$$\begin{aligned} &M_x w = \left(w(x,y) + w(x+1,y) \right)/2 \\ &\text{and} \quad M_y w = \left(w(x,y) + w(x,y+1) \right)/2, \end{aligned} \quad (10)$$

In the second step, the desired image is reconstructed by fitting the normal vector of the desired image to the estimated normal vector $\mathbf{n}^*$ with constraints from data fidelity. This step is called as image reconstruction (IR) step in the following. Mathematically, the result can be obtained by solving the following minimization problem:

$$\min_{\mu} \int_{\Omega} \left(|\nabla f| - \nabla f \cdot \frac{n}{|n|} \right) d\mu \quad \textit{subject to} \quad |P - A\mu| \le \varepsilon \quad (11)$$

where P is the acquired projection data and A represents the system transfer matrix, which depends on the projection geometry

However, it is numerically difficult to solve such problem by using the AL method directly. Thus, the second step is decomposed into two parts and solved separately. The first part of Eq. (11) is for normal vector fitting which can be solved by Euler-Lagrange (EL) method [6]. The updating function is:

$$\mu^{n+1} = \mu^n + \Delta t_2 \left(D_x^- \left(\frac{D_x^+ f^n}{T_3^n} - n_1 \right) + D_y^- \left(\frac{D_y^+ f^n}{T_4^n} - n_2 \right) \right), \quad (12)$$

where

$$T_3 = \sqrt{D_x^+ f^n + \left(M_x \left(C_y^1 f^n \right) \right)^2 + \varepsilon}, \quad (13)$$

$$T_4 = \sqrt{D_y^+ f^n + \left(M_y \left(C_x^1 f^n \right) \right)^2 + \varepsilon}, \quad (14)$$

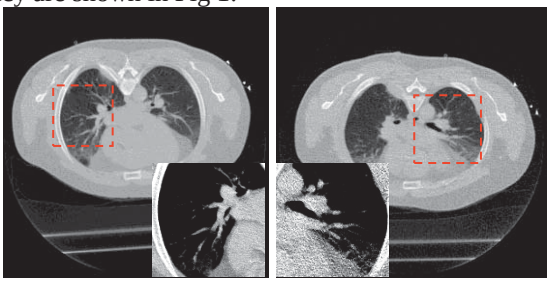
$$\begin{aligned} &n_1 = \frac{u}{\sqrt{u^2 + \left(M_x \left(M_y v \right) \right)^2 + \varepsilon}}, \\ &\text{and} \quad n_2 = \frac{v}{\sqrt{v^2 + \left(M_y \left(M_x u \right) \right)^2 + \varepsilon}} \end{aligned} \quad (15)$$

Δt_2 is another relaxing parameter to control the updating step length in the IR step. The constraints part in Eq. (11) can be solved by POCS strategy [1]. Therefore, the desired image can be obtained by solving the two minimization problem (i.e., TFS and IR) iteratively.

III. Results

In this study, to analyze the proposed method, two thorax CT raw data sets (i.e., normal-mAs scan and low-mAs scan) were acquired from a clinical multi-slice CT scanner in non-FFS model (Stony Brook Hospital, Stony brook, NY). The number of channels in each detector row is 672, the fan angle increment for each channel is 0.0775862° and the bin size along the z axis is 0.75mm. The tube angle increment is

0.3103448° (i.e., 1160 projection views per 360°). The radius of the focal spot circle is 570mm, and distance between the source and the detector plane is 1040mm. The tube voltage was set to be 120kVp and tube current was set to be 100mAs for normal-mAs scan and 20mAs for low-mAs scan. The field of view (FOV) is 51.2×51.2 cm^2, with corresponding pixel size is 1×1 mm^2. For the spiral CT system, the raw data is often rebinned to fan beam projection data for research purpose by considering the effects of pitch (i.e., the movement of the patient along the z axis). In our study, we use the rebinning method as introduced in [8] to rebin our spiral projection data to multi-slice fan beam sinogram data. In this patient's case, about 264 slices of sinogram data were obtained for normal-mAs scan and 139 slices for low-mAs scan for clinical purpose. In this study, we choose the 60th slice from normal-mAs and 91st slice from low-mAs full-views projection data for analysis. The transverse images of the projection data by FBP method with Hanning window at Nyquist cut-off frequency are shown in Fig 1.



(a) (b)

Figure 1: Thorax image reconstructions from full-view projection data acquired at (a) 100mAs and (b) 20mAs by FBP method. The display window is [0, 0.03] mm^{-1}. : The ROIs of the reconstructed images of the lung is displayed by the window [0.011, 0.023] mm^{-1}.

A. Normal-mAs case

Visualization-based evaluation

In this section, 116 and 58 projection views are extracted evenly from the rebinned 100-mAs sinogram data. Thus, the associative dosage are reduced to about 1/10 and 1/20 of the normal dosage, respectively. The reconstructed results are shown in Fig. 2. Regions of interests (ROIs) as indicated in Fig. 1(a) were selected to examine the fine structure of the reconstructed images in a small display window. The corresponding zoomed ROIs are shown in Fig. 3.

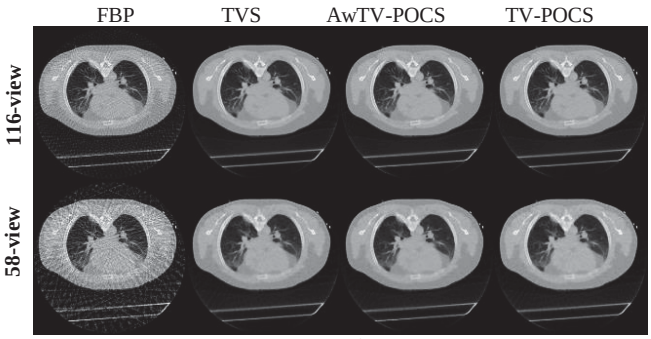


Fig. 2: The reconstructed images of the lung in 100mAs. The display window is [0, 0.03] mm^{-1}.

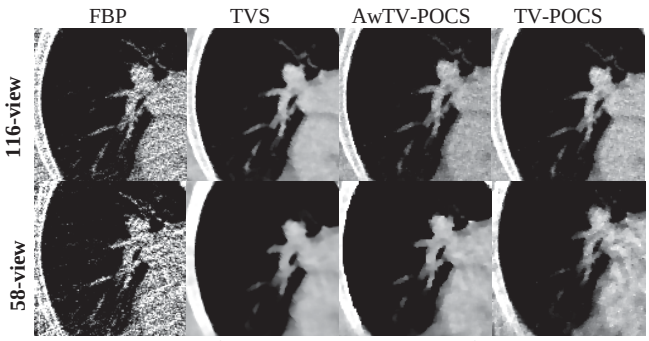


Fig. 3: The ROIs of the reconstructed images of the lung in 100mAs. The display window is [0.011, 0.023] mm^{-1}.

From figure 2, we can observe that in 116-view case, all the three methods can produce satisfied results when displayed in a large display window. However, in Fig. 3, the results from AwTV-POCS and TV-POCS show some small patch artifacts in the uniform area when examining in a small display window which are consistent with previous report [4]. In the contrary, the results from TVS method show smoothed images in the uniform area, which is consistent with the results in [6]. In addition, in the 58-view case, the image from TV-POCS shows undesired larger patch artifacts in the uniform area. Although the AwTV-POCS method shows ability to mitigate such patchy artifacts, the TVS method is more efficiently as shown in Fig. 3.

Universal quality index (UQI) measurement study

The universal quality index (UQI) is one of the image quantitative merits to measure the similarity between the reconstructed image and the reference image. Three factors: loss of correlation, luminance distortion and contrast distortion are considered in the UQI [3]. In this study, we selected a ROI as indicated in Fig 1(a) to exam the UQI value. It should be pointed out that the image f_0 reconstructed from normal-dose full-view data by FBP method yields a desired baseline and is used as reference image due to the lack of true image. Although different reference images could lead to different evaluation results, the image obtained from high-dose full-view data by FBP method has been generally accepted as a baseline for low-dose CT image reconstruction evaluation. We use f_1 to denote the testing images.

For N' pixels ROIs, the means, variances and covariances are defined as [3]:

$$\bar{f}_j = \frac{1}{N'}\sum_{n=1}^{N'} f_{jn}, \tag{17}$$

$$\sigma_j^2 = \frac{1}{N'-1}\sum_{n=1}^{N'}\left(f_{jn}-\bar{f}_j\right)^2, \tag{18}$$

where j=0 and 1, and

$$Cov\{f_1, f_0\} = \frac{1}{N'-1}\sum_{n=1}^{N'}\left(f_{1n}-\bar{f}_1\right)\left(f_{0n}-\bar{f}_0\right), \tag{19}$$

Then, the UQI is defined as:

$$UQI = \frac{2Cov\{f_1, f_0\}}{\sigma_1^2+\sigma_0^2}\frac{2\bar{f}_1\bar{f}_0}{\bar{f}_1^2+\bar{f}_0^2}, \tag{20}$$

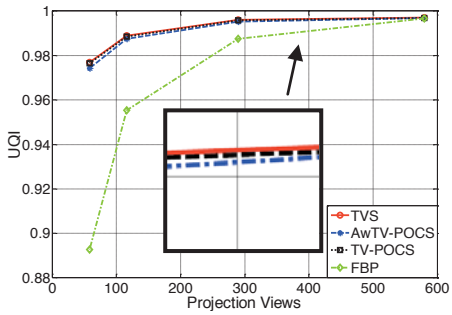


Figure 4: The UQI versus projection views curves from the lung image study.

The curve of UQI values within the selected ROIs versus the projection views (i.e., 58, 116, 290 and 580-view) by different methods (i.e., TVS, AwTV/TV-POCS and FBP) are shown in Fig. 4. From figure 4 we can observe that the results from TVS method have the highest UQI values and the results from AwTV-POCS method have the lowest UQI values. It should be mentioned that due to the weights for edge preserving in the AwTV model, which didn't considered in TV-POCS and TVS methods, the results of AwTV-POCS have a lower similarity to the results from normal-mAs full-view data by FBP, which only used spatially invariant filtering.

B. Low-mAs case

In this section, 116 and 58 projection views are extracted evenly from the low-mAs (i.e., 20mAs) projection data. Compared to the low-dose scan, the associative dosage are reduced to about 1/10 and 1/20 of the original dosage, respectively. The reconstructed results are shown in Fig. 5. A ROI as indicated in Fig. 1(b) was selected to examine details of the reconstructed image. The corresponding ROIs results are shown in Fig. 6.

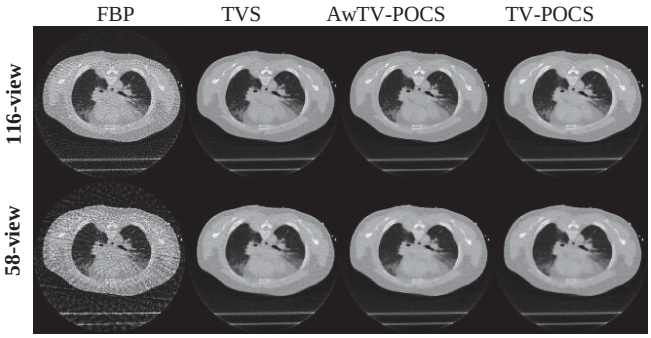


Fig. 5: The reconstructed images of the lung in 20mAs. The display window is [0, 0.03] mm^{-1}.

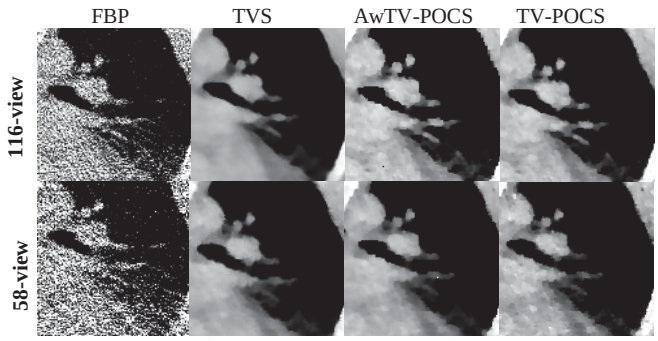


Fig. 6: The ROIs of the reconstructed images of the lung in 20mAs. The display window is [0.011, 0.023] mm^{-1}.

In figure 5 and 6, the images show more artifacts compared to the results from normal-dose case due to the inconsistent of the projection data. In addition, more patchy artifacts are observed in the results of AwTV/TV-POCS methods as the number of projection decreases. The results from TVS methods show fewer artifacts in the uniform area, which is consistent with our previous observation in normal-dose case.

IV. Conclusion and Discussion

In this work, we presented the TVS method for low-dose image reconstruction from sparse-view data. This method is a new attempt to eliminate the effects of staircase and patchy artifacts caused by the piecewise constant assumption. The preliminary results indicate that the TVS method has advantages over the AwTV/TV-POCS methods at the uniform region for image reconstruction from sparse-view data. In addition, although it is numerically hard to prove the edge preserving property of the TVS, the results suggest that the TVS method can preserve edges quite well.

In our preliminary study on the sparse-view cases, we observe some patchy artifacts in the uniform area from the results of AwTV/TV-POCS methods which follows a piecewise constant assumption. However, through this TVS strategy, the undesired patchy artifacts are eliminated efficiently due to the consistency along both of the normal direction and tangent direction. In our future work, we will focus on the evaluations of this new TVS method using more simulated and experimental data.

References

[1] E. Sidky and X. Pan, "Image reconstruction in circular cone-beam CT by constrained, total-variation minimization", *Physics in Medicine and Biology*, vol. 53, pp. 4777–4807, August, 2008.

[2] G. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS)", *Proc. Soc. Photon Instrum. Eng*., vol. 6856, pp. 1–18, March, 2008.

[3] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. a. Pelizzari and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Physics in Medicine and Biology*, vol. 55, pp. 6575-6898, October, 2010.

[4] J. Tang, B. E. Nett and G. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Physics in Medicine and Biology*, vol. 54, pp. 5781-5804, October, 2009.

[5] Y. Liu, J. Ma, Y. Fan and Z. Liang, "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," *Physics in Medicine and Biology*, vol. 57, pp. 7923-7956, November, 2012.

[6] T. Rahman, X. Tai, and S. Osher, "A TV-stokes denoising algorithm", SSVM'07 Proceedings of the 1st international conference on Scale space and variational methods in computer vision, pp. 473-483, 2007.

[7] X. Tai, S. Osher and R. Holm, " Image Inpainting Using a TV-Stokes Equation", *Image Processing Based on Partial Differential Equations Mathematics and Visualization*, pp.3-22, 2007.

[8] F. Noo, M. Defrise and R. Clackdoyle, "Single-slice rebinning method for helical cone-beam CT," *Physics in Medicine and Biology*, vol. 44, pp. 561-570, February, 1999.

Exploitation of geometric a priori knowledge for limited data reconstruction in non-destructive testing

Christian Schorr, Michael Maisl

Abstract—**Computed tomography (CT) is a very powerful tool in medicine and non-destructive testing but is unsuitable for planar objects. A solution can be found in the use of computed laminography (CL), a technique where the object is irradiated by an oblique angle thereby circumventing the problems arising in CT. Due to the limited amount of angular coverage and the special geometric set-up, filtered backprojection methods [4] cannot be employed for the reconstruction in this case. More flexible iterative algorithms like SART (simultaneous reconstruction technique) [1,2] provide an answer to this challenge. One of their important advantages when compared to filtered backprojection methods is their ability to incorporate a priori information about the object into the reconstruction process [6]. Often the object's geometry is known from CAD files or other technical specifications. Especially in the case of limited-angle data, where only a part of the object can be measured, and laminographic geometries, additional information is of great importance. This geometrical a priori knowledge can be exploited to restrict the reconstruction volume to areas where material is definitely present, resulting in correct object contours even in the limited-angle case. This reduces artifacts and increases contrast thereby allowing for a better defect detectability and thus an easier and more reliable inspection of the object.**

***Index Terms*— A priori knowledge, computed laminography, computed tomography, iterative reconstruction, limited data**

I. Introduction

Computed tomography (CT) is a well-established and widely used non-destructive inspection method for inspecting the interior structure of objects. Applying standard reconstruction methods for circular or helical sampling to planar objects, two fundamental problems arise: impenetrability in longitudinal direction and collision risks between X-ray source and object at high magnifications. During a CT, the object is rotated by 360 degrees while being irradiated. Planar objects are challenging since they exhibit very different irradiation lengths. In normal direction to the surface absorption is very much lower than in longitudinal direction. Trying to compensate for this by increasing the energy of the X-rays, one automatically reduces contrast and geometrical resolution, thereby possibly rendering the reconstruction useless. The opening angle of the X-ray source allows for a variation of magnification by changing the distance between X-ray source and object. Small object features can be inspected in detail this way. Especially planar objects with very fine structures can require such a high magnification, that the required source-detector distance gets too small to allow for a full 360° rotation without risking a collision between source and object. Circumventing this problem by increasing both source-object and source-detector distances while keeping the desired magnification ratio results in a severely limited opening angle which in turn restricts the field of view making multiple scans necessary to cover the entire area of interest. Computed laminography (CL) can solve these problems. In contrast to standard CT geometries where X-ray source and detector are perpendicular to each other and the axis of rotation and a full 360° coverage is neccessary , CL can also work with a limited angular range of less than 90° (Swing laminography) or completely dispense with the traditional set-up and use linear translational (translation CL), planar rotational (classic CL) or tilted geometries (CLARA (Computed Laminography And RAdiography)) [7,8]. The advantage of all these trajectories lies in their possibility to place the object close enough to the source to achieve the desired resolution without colliding with the X-ray tube. Additionally most of these geometries allow for a constant oblique irradiation angle throughout the entire measurement. This eliminates the problem of widely differing object thicknesses with all the drawbacks mentioned above.

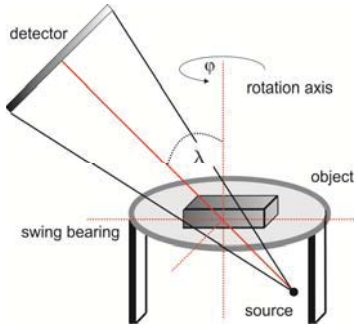


Fig. 1. CLARA geometry, the object is situated on a swing bearing, thus allowing the source beneath to irradiate it by an oblique laminography angle lambda. The detector is arrayed perpendicular to the central ray of the source. For the measurement, the object is rotated 360° around the rotation axis.

Submitted to Fully3D conference on 4th February 2013
C. Schorr is with the Fraunhofer Institute for Non-destructive Testing (IZFP), Campus E 3.1, 66123 Saarbruecken, Germany (e-mail: christian.schorr@ izfp.fraunhofer.de).

M. Maisl, PhD is with the Fraunhofer Institute for Non-destructive Testing (IZFP), Campus E 3.1, 66123 Saarbruecken, Germany (e-mail: michael.maisl@ izfp.fraunhofer.de).

II. MATHEMATICAL BACKGROUND

The SART algorithm is ideally suited to incorporate geometric a priori knowledge. It computes the reconstruction of the density distribution of an object by iteratively solving a system of linear equations. The physical process of a CT measurement is modelled as a matrix-vector equation and solved iteratively employing the Kaczmarz algorithm [5]:

Be v a volume consisting of N cubic voxels j with constant values v_j. Furthermore be p the vector of measured rays of dimension M with p_i the ray sum along the i. ray passing through v. Then the relation between v and p can be expressed by the following equation:

$$Wv = p,\ \sum_{j=1}^{N} w_{ij} v_j = p_i, i = 1, 2, \dots, m \tag{1}$$

where W is a matrix of dimension M x N with entries w_{ij} representing the weighting coefficient in voxel j along ray i. w_{ij} can be interpreted as the contribution of the entire ray p_i lying in voxel j. Be v_j^k the value of voxel j after k projections, λ a relaxation factor and P the projection belonging to angle φ. Then the SART algorithm is defined as:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{\sum_{p_i \in P_\varphi}^{i} \left(\frac{p_i - \sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in} v_n^{(k)}}{\sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in}} \right) w_{ij}}{\sum_{p_i \in P_\varphi}{}_{i} w_{ij}} \tag{2}$$

Often knowledge about the object to be inspected is available before the measurement. This a priori information usually consists of geometric information about the form of the object and its dimensions. Exploiting a priori knowledge can help to improve the quality and the convergence speed of the reconstruction. Geometrical a priori information about the contours of the object is implemented by an additional volume of the same dimensions as the volume to be reconstructed. This weighting volume G contains voxels of value $g_j \in \{0; 1\}$, signifying the voxels' probability to contain either nothing (air) or material. A direct approach [3] would then be to use these g_j as an additional weighting for the back-projection, resulting in the API-SART algorithm:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{\sum_{p_i \in P_\varphi}{}_{i} \left(\frac{p_i - \sum_{n=1}^{N} w_{in} v_n^{(k)}}{\sum_{n=1}^{N} w_{in}} \right) w_{ij} g_j}{\sum_{p_i \in P_{\varphi_i}}{}_{i} w_{ij}} \tag{3}$$

A drawback of this method is the loss of „material“ since the additive correction values c_j assigned to voxels a priori weighted by 0 during back-projection are lost.

Using the following ray length correction process [9] alleviates this problem: After computing the voxels j, into which will be back-projected and their ray lengths w_{ij} , the corresponding weights g_j which satisfy $g_j = 0$ are determined. The computation of the corrected overall ray length takes only those voxels j into account whose corresponding $g_j \neq 0$. With this adapted overall ray length the weights used in the back projection change since the overall ray length has been diminished which in turn increases the values of w_{ij}. This concept leads to a ray-length-corrected weighting g_j^{RLC} for the backprojection:

$$g_j^{RLC} = \left(\frac{\sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in}}{\sum_{n=1}^{N}{}_{g_n \neq 1} w_{in}} \right) g_j,\ w_{in}, g_j \in \mathbb{R} \tag{4}$$

Substituting g_j in the basic a priori SART algorithm (3) by the ray-length-corrected g_j^{RLC} from (4) leads to the ray-length-corrected API-SART method:

$$v_j^{(k+1)} = v_j^{(k)} + \lambda \frac{\sum_{p_i \in P_\varphi}{}_{i} \left(\frac{p_i - \sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in} v_n^{(k)}}{\sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in}} \right) w_{ij} \left(\frac{\sum_{n=1}^{N}{}_{w_{in} \neq 0} w_{in}}{\sum_{n=1}^{N}{}_{g_n \neq 1} w_{in}} \right) g_j}{\sum_{p_i \in P_\varphi}{}_{i} w_{ij}} \tag{5}$$

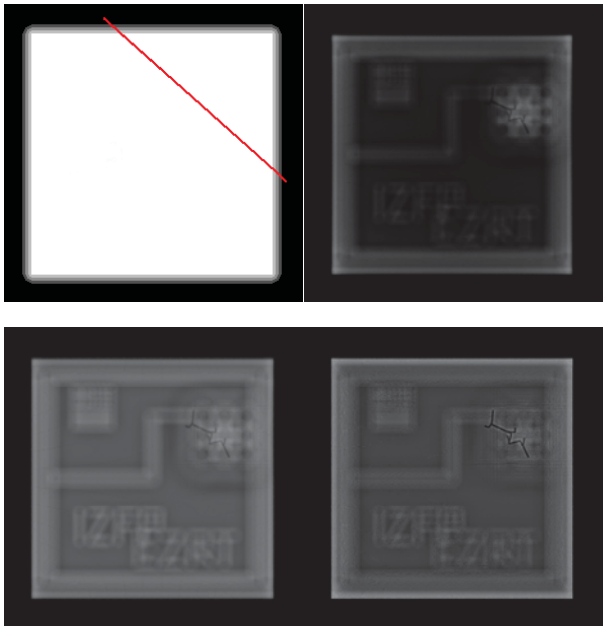
III. EXPERIMENTAL RESULTS

The proposed algorithm was tested both on simulated and real data. For the CLARA geometry a simulated printed circuit board (PCB) phantom with a crack was generated (Fig. 6 (a)). To investigate the algorithm's performance on real objects, a ski boot buckle made of a magnesium alloy and containing porosities of different shapes and sizes was scanned with a limited-angle CT of 90° (Fig. 6 (b)).

The PCB-phantom was simulated using the software Scorpius XLab. A simulated 360° CT reconstruction of the phantom without crack computed from 400 projections served as a starting point for the generation of the a priori weighting volume. The reconstruction was binarized, filling in all interior defects to guarantee an a priori volume containing only the outer contours of the object. Three successive layers of 5 voxels thickness each with a priori weights 0.75, 0.5 and 0.25 respectively were applied to the boundaries of the binarized material area inside the volume (see fig. and). A CLARA geometry measurement of 40 projections and $\lambda = 45°$ with crack was then simulated and reconstructed using the a priori volume from the CT simulation.

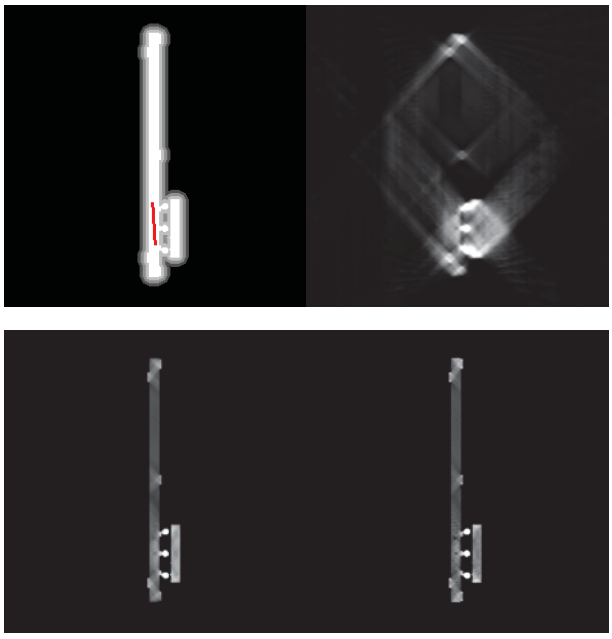
For the ski buckle, 800 projections were taken using a CT scanner equipped with a Perkin-Elmer XRD 1620 CN detector of 2048^2 pixels and a Viscom XT 9225 D X-ray tube. From these data a CT reconstruction was computed and binarized. Defects in the resulting volume were then eliminated and the same layering process as described above was applied. A 90° range of 100 projections taken from the full data set was used in conjunction with the a priori volume to reconstruct the buckle.

All reconstructions were computed using 3 iterations of the respective algorithms with a resolution of 512^3 voxels. The contrast for each picture was chosen differently to deliver the most suitable result for optical inspection by the user, as is usual in NDT practice.



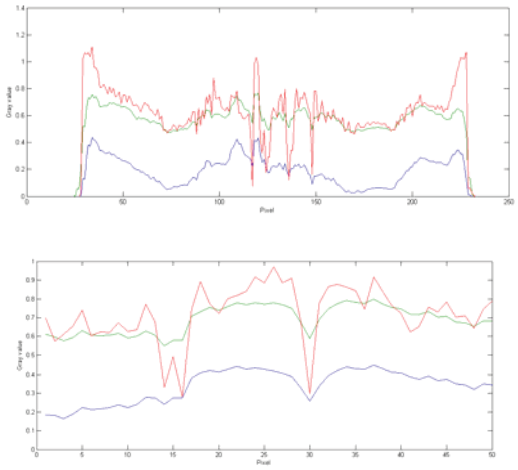
(c) (d)

Fig. 2. PCB Phantom, frontal view - (a) a priori volume, (b) SART reconstruction without a priori information, (c) SART reconstruction with a priori information, (d) SART reconstruction with a priori information and ray-length-correction



(c) (d)

Fig. 3. PCB Phantom, side view - (a) a priori volume, (b) SART reconstruction without a priori information, (c) SART reconstruction with a priori information, (d) SART reconstruction with a priori information and ray-length-correction

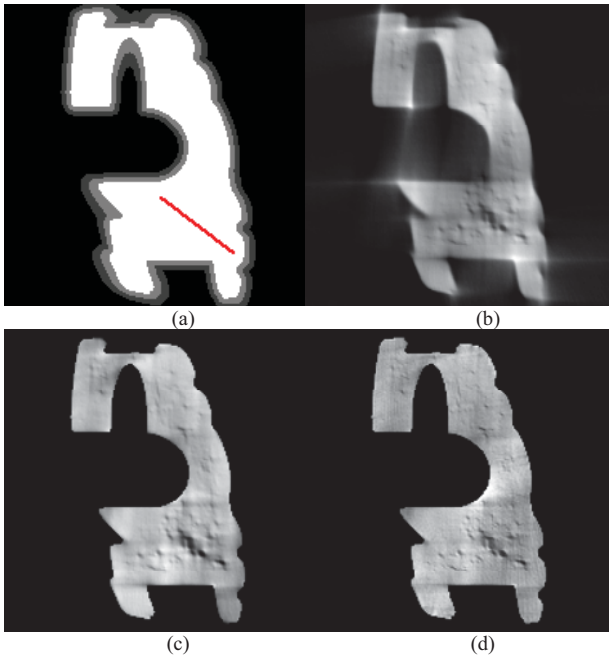


(b)

Fig. 4. PCB phantom, (a) - gray value profile along red line in fig. 3 (a), (b) - gray value profile along red line in fig. 4 (a), SART reconstruction without a priori information (blue), SART reconstruction with a priori information (green), SART reconstruction with a priori information and ray-length-correction (red)

When comparing the reconstructions of the PCB phantom, the effect of the additional a priori weighting is obvious. The frontal (Fig. 2) view clearly shows the crack in the PCB if a priori methods are used (c) and especially if an additional ray-length-correction (d) is applied. A standard SART reconstruction (b) suffers from poor contrast which renders the crack hard to detect. The side view (Fig. 3) of the SART reconstruction (b) offers the typical smeared-out contours of a laminographic set-up with $\lambda = 45°$ leaving only one branch of the crack just beneath the middle ball visible. With a priori information (c) the contours are sharp, but the second branch of the crack is only detectable if the reconstruction is also ray-length-corrected (d). Fig. 4 shows the gray value profiles through the reconstruction along the red lines in Fig. 3 (a) and Fig. 4 (a). The improved a priori SART algorithm with ray-length-correction significantly increases contrast compared to both a standard SART and a priori SART algorithm. This can be seen in Fig. 4 (a) between pixels 120 and 150 and (b) around pixels 150 and 200, where the cracks are located. It also introduces a small amount of structured noise which is easily off-set by the increased defect detectability.

The frontal view (Fig. 5) of the standard SART reconstruction (b) of the ski boot buckle data set exhibits the effects of the 90° angular range restriction. There are streak artifacts along the objects corners and the interior semicircular cavity is reconstructed as a quarter circle only. Incorporating a priori information about the buckle's geometry results in a much better contour reconstruction (c). An additional ray-length-correction (d) renders the porosities in the lower right part with an increased contrast. Fig. 6 shows that the contrast between SART and a priori SART is almost equal, whereas the ray-length-corrected algorithm improves contrast significantly. Observe that there is also a slight increase in noise, as already seen in the PCB phantom.



(a) (b)

(c) (d)

Fig. 5. PCB Phantom, frontal - (a) a priori volume, (b) SART reconstruction without a priori information, (c) SART reconstruction with a priori information, (d) SART reconstruction with a priori information and ray-length-correction

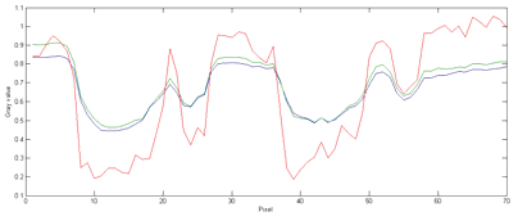
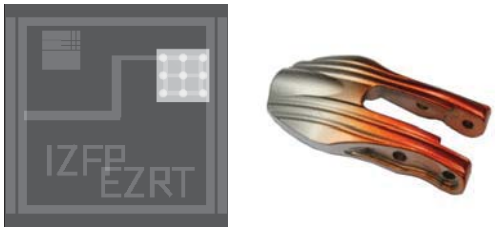


Fig. 6. Ski boot buckle, gray value profile along red line in fig. 5 (a), SART reconstruction without a priori information (blue), SART reconstruction with a priori information (green), SART reconstruction with a priori information and ray-length-correction (red)



(a) (b)

Fig. 6. Test objects - (a) PCB phantom, 3D visualization - (b) ski boot buckle

IV. CONCLUSION

The proposed a priori strategy is able to significantly improve the reconstruction of the object's contours. Introducing and additional ray-length-correction into the a priori SART algorithm greatly increases the contrast when compared with standard SART results. For laminographic measurements, where only an oblique irradiation angle is possible, and for limited-angle CT data, incorporating geometric a priori knowledge proves to be an promising approach to improve defect detectability in spite of the artifacts introduced by the limited data available. Test on both simulated and real measurements for CLARA and limited-angle CT geometries have shown the usefulness of the proposed a priori SART algorithm with ray-length correction. There is a certain amount of structured noise introduced by the ray-length correction but the increase in contrast by a factor of 2-3 more than compensates for this effect. If planar objects like PCBs or sheets of fibre reinforced composites are to be scanned, a priori enhanced SART algorithms combined with laminographic geometries are well suited for pratical use.

REFERENCES

[1] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm", *Ultrasonic Imaging*, vol. 6, no. 1, pp. *81-94,* 1984.

[2] P. Gilbert, "Iterative Methods for the Three-dimensional Reconstruction of an Object from Projections", *Journal of Theoretical Biology*, vol. *36,* no. 1, pp. *105-117,* 1972.

[3] S. Gondrom, "Rekonstruktion von Objektebenen aus Röntgendurchstrahlungsaufnahmen bei unvollständigen Daten unter Verwendung von a priori Information.", Ph.D. thesis, UDS, Saarland, Germany, 2001

[4] L. Feldkamp, L. Davis and J Kress, "Practical cone-beam algorithm", *Journal of the Optical Society of America A,* vol. 1, no. 6, pp. *612-619,* 1984.

[5] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen", *Bulletin International de l'Academic Polonaise des Sciences et des Lettres*, vol. 35, pp. *355-357,* 1937.

[6] L.B. Meyer and G.T. Herman, "Algebraic reconstruction techniques can be made computationally efficient", *IEE Transactions on medical imaging*, vol. *12,* no. 3, pp. *600-609,* 1993.

[7] F. Porsch, U. Hassler, M. Maisl and C. Schorr, "X-ray computed laminography: an approach of computed tomography for applications with limited access", in *Leitfaden zur industriellen Röntgentechnik: Zerstörungsfreie Prüfung mit Bildverarbeitung, 11*th ed. Erlangen.

[8] F. Porsch, C. Schorr and M. Maisl, "Computed laminography for x-ray inspection of lightweight constructions", *Deutsche Gesellschaft für Zerstörungsfreie Prüfung e.V. (DGZfP): International Symposium on NDT Aerospace*, Berlin, Germany, 2010

[9] C. Schorr and M. Maisl, "Computed laminography using a priori information.", Presented at Proceedings, *Conference on Industrial Computed Tomography (ICT),* Wels, Austria, 2012

Truncation Correction using a 3D Filter for Cone-beam CT

Yan Xia, Andreas Maier, Frank Dennerlein, and Joachim Hornegger

***Abstract*—Recently, a novel method for region of interest (ROI) reconstruction from truncated projections with neither the use of prior knowledge nor explicit extrapolation has been published, named Approximated Truncation Robust Algorithm for Computed Tomography (ATRACT). It was derived by analytically reformulating the standard Feldkamp-Davis-Kress (FDK) algorithm into a reconstruction scheme that is by construction less sensitive to lateral data truncation. In this paper, we present and investigate a variation of the ATRACT that is to apply ATRACT in 3D by decomposing the ramp filter into the 3D Laplace filter and a 3D residual filter. ROI reconstruction can be readily realized by performing these two successive filters on projection data stack at once and followed by standard backprojection. Real data evaluation shows that the new method at least performs as well as the native ATRACT in terms of truncation correction. However, for off-center reconstruction, the linear gradient artifact arose in native ATRACT is essentially reduced by the new method.**

I. Introduction

It is well known that the X-ray radiation dose exposed to the patient during a CT exam is proportional to the volume that is irradiated during the scan. Several medical applications require only a small volume to be imaged. For example, in the neurointerventional radiology only micro devices, e.g. implanted stents or coils, are required to be examined in multiple times. Although only the small area is of diagnostic interest, conventionally, a scan with a full field of view (FOV) was performed, resulting in a considerable dose to the patient. Hence, a restriction of the X-ray beam to only that area would significantly reduce radiation dose. This is simply done by deploying a collimator near the X-ray source. However, the resulting lateral truncation in projections, poses a challenge to the conventional tomographic reconstruction algorithms.

So far many algorithms specially concerning the ROI reconstruction have been proposed. Some are based on the requirement of prior knowledge on the reconstructed object so that the ROI problem can be exactly solved [1], [2], [3]. Other approaches estimate the missing data using an extrapolation procedure as a pre-processing step [4], [5], [6], [7].

A novel method (ATRACT) has been suggested for ROI reconstruction with neither the use of prior knowledge nor explicit extrapolation [8]. In this method, the standard ramp filter is decomposed into the 2D Laplace filtering and a 2D Radon-based filtering step or 2D convolution-based filtering. In this paper, we present and investigate a variation of the original ATRACT that is to apply ATRACT in 3D by decomposing the ramp filter into the 3D Laplace filter and a 3D residual filter. We expect the 3D convolution-based filtering will gain more stability than its 2D counterpart, especially in the truncated edge.

This paper is organized as follows. Section II reviews the ATRACT algorithm and presents new 3D ATRACT algorithm. Experiment setups are specified in section III, and reconstruction results from these setups are presented in section IV. The paper ends with conclusions, in section V.

Y. Xia, A. Maier and J. Hornegger are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany. Y. Xia and J. Hornegger are also with the Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-University Erlangen-Nuremberg, 91052 Erlangen, Germany. (e-mail: yan.xia@cs.fau.de; andreas.maier@cs.fau.de; joachim.hornegger@cs.fau.de).

F. Dennerlein is with Siemens AG, Healthcare Sector, 91052 Erlangen, Germany (e-mail: frank.dennerlein@siemens.com).

II. Truncation Correction Methods

A. 2D ATRACT

Intuitively, the idea behind ATRACT is to adapt the Feldkamp-Davis-Kress (FDK) algorithm [9] by decomposing the 1D ramp filter operation into two successive 2D filtering steps — the 2D Laplace filtering and a 2D Radon-based filtering step — one acting locally and one acting non-locally on the projection data. We refer to this method as 2D ATRACT in the following. In presence of lateral data truncation, 2D ATRACT allows us to exclude the artifacts typically occurring during filtering, simply by removing the singularities (spikes) at the edges of lateral data truncation after the Laplace operation. With the FDK method, such a removal is not straight-forward, due to the non-local character of the ramp filter. In its later version, the Radon-based filter was substituted by a 2D convolution-based filter for increasing computational performance [10], [11]. This naturally inspires the idea of an alternative decomposition of the 1D ramp filter in 3D convolutions, to further improve the image quality.

B. 3D ATRACT

As discussed above, the 3D ATRACT algorithm is also obtained by a modification of the standard ramp filter in FDK algorithm. That is to decompose the ramp filter into the 3D Laplace filter and a 3D residual filter.

Fig. 1 shows the associated notations in the cone-beam short-scan imaging geometry. The mathematical expression of 3D projection data stack $g(\lambda, u, v)$ can be written as follows:

$$g(\lambda, u, v) = \int_0^\infty f(\mathbf{a}(\lambda) + t\boldsymbol{\alpha}(\lambda, u, v))\,\mathrm{d}t\,, \qquad (1)$$

where u, v are flat detector coordinates and λ indicates angular coordinate.

Using the notations that are shown in Fig. 1, the 3D ATRACT algorithm can be written as follows:

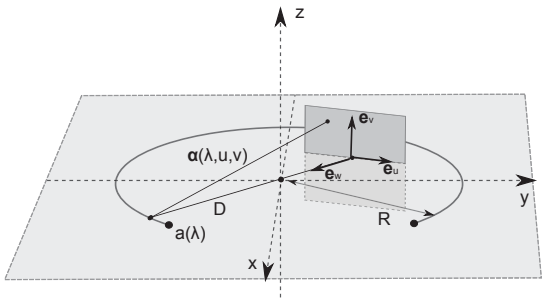


Fig. 1. Cone-beam geometry and associated notation: The curve $\mathbf{a}(\lambda) = (R\cos\lambda, R\sin\lambda, 0)$ describes the trajectory of the X-ray source, with the scan radius R and the rotation angle λ. The planar detector is parallel to the unit vectors $\mathbf{e_u}(\lambda)$ and $\mathbf{e_v}(\lambda)$ and at distance D from the source. $\mathbf{e_w}(\lambda)$ is the detector normal. We use the 3D function $g(\lambda, u, v)$ to describe the projection data stack at the point (u, v) acquired at angle λ.

Step 1: Cosine- and Parker-like weighting of projection data to obtain pre-scaled projection data $g_1(\lambda, u, v)$:

$$g_1(\lambda, u, v) = \frac{Dm(\lambda, u)}{\sqrt{D^2 + u^2 + v^2}} g(\lambda, u, v) \tag{2}$$

where $m(\lambda, u)$ is Parker weight for short-scan data.

Step 2: 3D Laplace filtering to obtain projection data $g_2(\lambda, u, v)$:

$$g_2(\lambda, u, v) = \left(\frac{\partial^2}{\partial\lambda^2} + \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right) g_1(\lambda, u, v) \tag{3}$$

Step 3: 3D convolution-based residual filtering to get filtered projection data $g_F(\lambda, u, v)$:

$$g_F(\lambda, u, v) = \int_{u_1}^{u_2}\int_{v_1}^{v_2}\int_{\lambda_1}^{\lambda_2} g_2(\lambda - \lambda', u - u', v - v') \, h_{3D}(\lambda', u', v')\, \mathrm{d}\lambda' \mathrm{d}u' \mathrm{d}v' \tag{4}$$

Step 4: 3D cone-beam backprojection to get the estimated object function $f^{(ATRACT)}(x, y, z)$:

$$f^{(ATRACT)}(x, y, z) = \int_{\lambda_1}^{\lambda_2} \frac{RD}{\left[R - \mathrm{x}\cdot\mathbf{e}_w(\lambda)\right]^2} g_F(\lambda, u, v)\mathrm{d}\lambda \tag{5}$$

where $\mathrm{x} = (x, y, z)$.

As illustrated in Fig. 2, reconstructions from the truncated data can be readily realized by performing two successive 3D filters on pre-scaled 3D projection data stack at once and followed by standard backprojection. Similarly, 3D ATRACT is able to exclude artificial high frequencies (removal of high spikes) after the 3D Laplace filtering step. Unlike the 2D ATRACT, additional removal of the spikes in λ direction is required. That means either removal of the first and last projections or constantly extrapolate them to avoid abrupt changes. Subsequently, the 3D residual filtering is carried out to obtain the desired filtered projections. In practice, the 3D Laplace operation can be achieved using a $3 \times 3 \times 3$ kernel with a different angluar weighting in spatial domain. The 3D residual filtering can be implemented by using 3D FFT-based convolution, and the residual kernel in Fourier domain is given by:

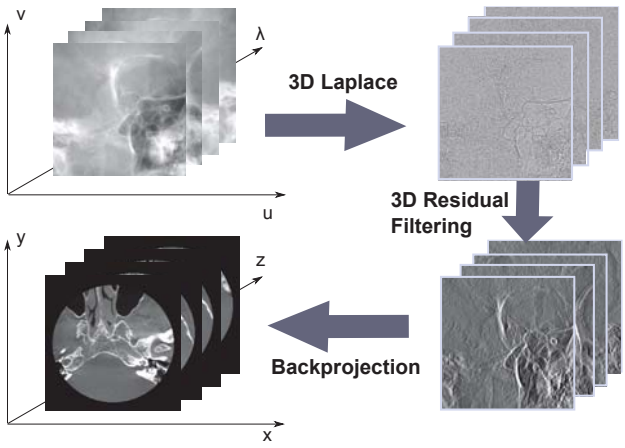


Fig. 2. Illustration of the 3D ATRACT algorithm. ROI reconstruction can be readily realized by performing two successive filters on the data stack at once and followed by the standard backprojection.

$$H_{3D}(\omega_\lambda, \omega_u, \omega_v) = -\frac{|\omega_u|}{\omega_\lambda^2 + \omega_u^2 + \omega_v^2} \tag{6}$$

Also note that discretization in u and v (i.e. $\mathrm{d}u$ and $\mathrm{d}v$) is identical (square pixels assumed) while discretization in λ differs a lot. Thus, in numerical implementation we choose a different discretization in λ, i.e. $\mu\mathrm{d}\lambda$, where $\mu = 0.025$ is a scaling factor.

III. Experiment Setup

The proposed algorithm was evaluated by the following datasets in terms of spatial resolution, low contrast resolution as well as robustness of correction quality. All datasets are acquired on a C-arm system (Siemens AG, Healthcare Sector, Forchheim, Germany) and contain 496 projection images (1240×960) with effective pixel size of 0.308×0.308 mm^2 in 2×2 binning mode.

To evaluate the spatial resolution and low contrast resolution of the reconstructions from the new algorithm, we used a Siemens cone-beam phantom that contains several low- and high-contrast inserts useful for evaluation of image quality. We also used two clinical datasets acquired from St. Lukes' Episcopal Hospital (Houston, TX, USA), to quantify the robustness of the truncation correction in practical application.

In the following evaluation, two scenarios were considered. In the baseline scenario, no collimation was applied during the scan, yielding non-truncated projections on the entire area of the detector. In second scenario, we virtually cropped projection images so that only the small region of interest was kept. The non-truncated projections were reconstructed by FDK, which was used as the reference here. The virtually truncated projections, in which only up to 30% of the FOV remained compared to non-truncated projections, were reconstructed by the new algorithm. We also investigated the performance of the 2D ATRACT method, and compared it to the new correction method.

Analogous to the 2D ATRACT algorithm, the new algorithm also suffers from a global volume scaling artifact. A correction of scaling and bias was performed to align the range of values between FDK and the new method.

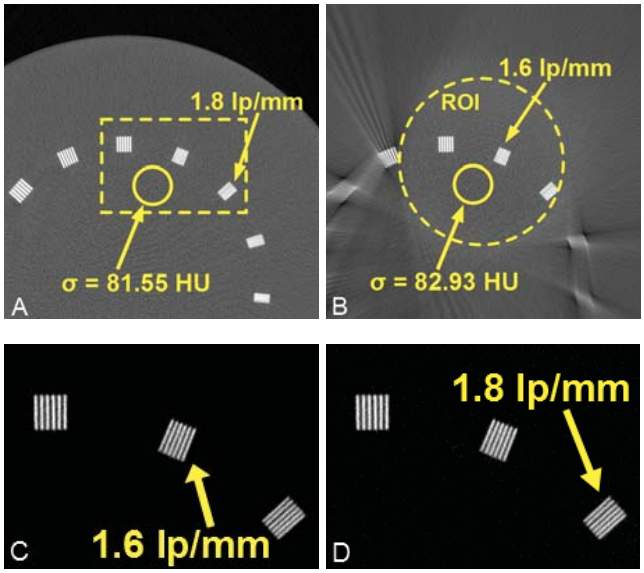


Fig. 3. Reconstruction results of the line-pair inserts phantom. Slice thickness is 0.25mm. A) and C): Standard FDK reconstruction from non-truncated projections, B) and D): 3D ATRACT reconstruction from virtually truncated projections.

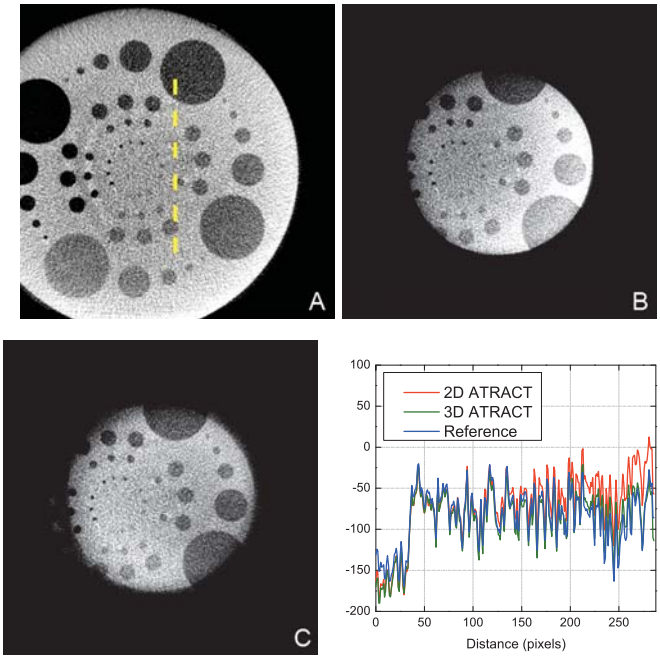


Fig. 4. Reconstruction results of the low contrast inserts in the gray scale window [-200HU, 0HU]. Slice thickness is 0.3mm. A): Standard FDK reconstruction from non-truncated projections, B) and C): 2D ATRACT and 3D ATRACT reconstruction from virtually truncated projections. Line profiles along yellow-dashed line in all methods are also provided.

IV. Results

A. Spatial Resolution and Low Contrast Resolution

Fig. 3 shows the reconstructions of the line-pair phantom. The investigated line-pair inserts (shown in yellow dashed box) in a clockwise direction have modulation of 1.4 lp/mm, 1.6 lp/mm and 1.8 lp/mm, respectively. The noise level of the given slices, estimated by computing the standard deviation within the yellow cycles, is 81.55 HU for the standard FDK reconstruction, and 82.93 HU for the 3D ATRACT-based ROI reconstruction. The reconstruction results confirm that 3D ATRACT reconstruction yields, for the investigated inserts, identical spatial resolution to the full FOV reconstruction by FDK.

Reconstructions of the low contrast inserts from FDK, 2D ATRACT and 3D ATRACT are represented in Fig. 4. The line profile along the yellow-dashed line in each slices is also given in right bottom. No significant differences are observed between ROI reconstructions and the reference reconstruction in terms of low contrast resolution. We found the result from

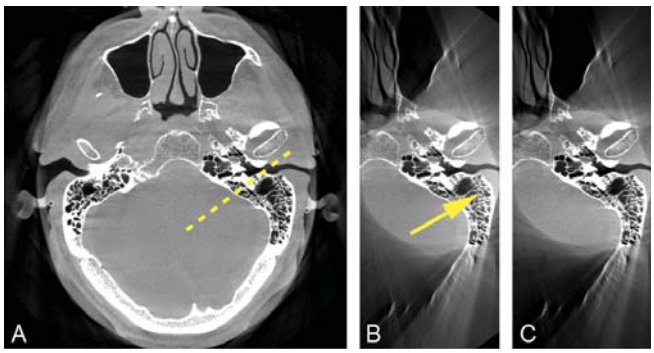


Fig. 6. Transversal slices of the clinical dataset 2 by the three algorithms, in the grayscale window [-1000HU, 1000HU]. Slice thickness is 0.4mm. A): Standard FDK reconstruction from non-truncated projection, B): 2D ATRACT-based ROI reconstruction, C): 3D ATRACT-based ROI reconstruction.

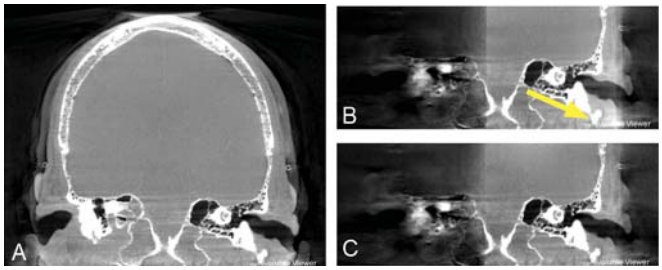


Fig. 7. Coronal slices of the clinical dataset 2 by the three algorithms, in the grayscale window [-1000HU, 1000HU]. Slice thickness is 0.4mm. A): Standard FDK reconstruction from non-truncated projection, B): 2D ATRACT-based ROI reconstruction, C): 3D ATRACT-based ROI reconstruction.

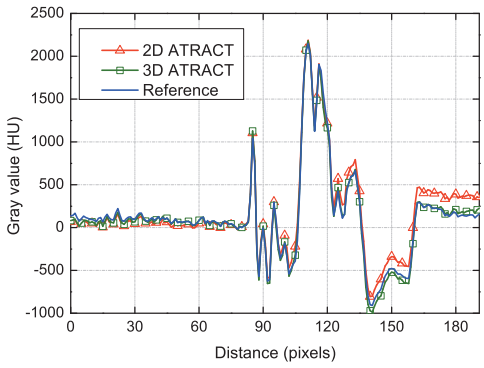


Fig. 8. Profiles along the yellow-dashed line shown in the transversal slices of clinical dataset 2.

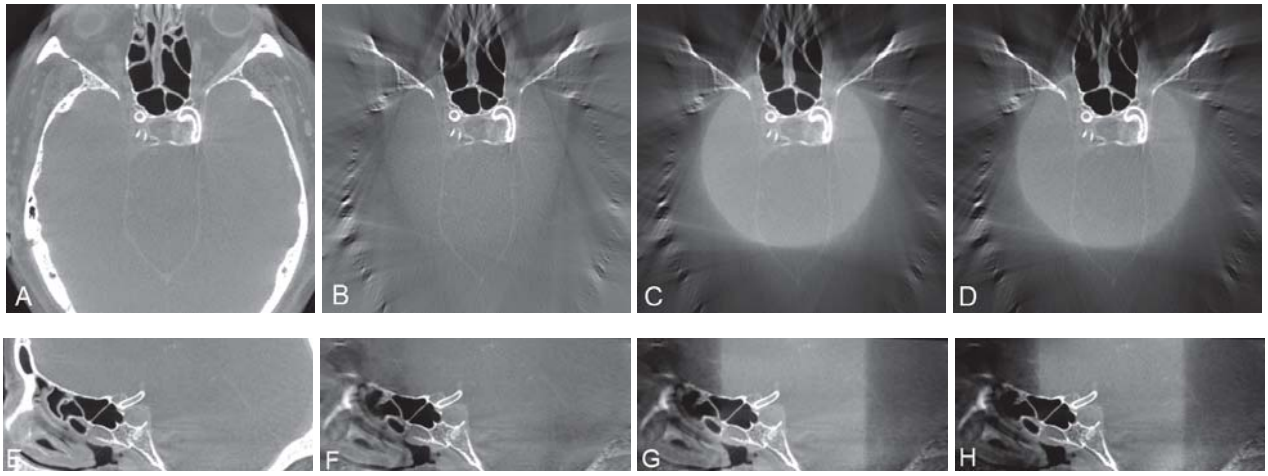


Fig. 5. Reconstruction results of the clinical dataset 1 by the three algorithms, in the grayscale window [-1000HU, 1000HU]. Slice thickness is 0.35mm. A) and E): FDK reconstruction from non-truncated projection, B) and F): constantly extrapolated FDK-based ROI reconstruction, C) and G): 2D ATRACT-based ROI reconstruction, D) and H): 3D ATRACT-based ROI reconstruction. The regions outside the field-of-view are masked in black.

2D ATRACT avoids the cupping artifact, but comes with a small linear gradient due to off-center reconstruction. Note that such artifacts were also observed in previous work [10]. A better result is obtained by 3D ATRACT that yields a reconstruction close to the reference.

B. Correction Quality

Reconstruction results of the clinical dataset 1 are shown in Fig. 5. It is clear that the straightforward FDK algorithm with a constant extrapolation cannot completely avoid the radial gradient-like truncation artifacts. As opposed to FDK-based ROI reconstruction, satisfying results are obtained by the proposed method and 2D ATRACT. No radial artifacts in the FOV are observed, which implies that truncation artifacts are essentially suppressed by the two methods.

In the clinical dataset 2, we deliberately applied an asymmetric collimation and thus resulted in the off-center ROI reconstruction. Transversal slices are represented in Fig. 6 and coronal slices are in Fig. 7. We observed that the overall correction quality in the reconstruction by both methods is maintained. However, it is noted that intensities tend to increase near the outmost edges of the 2D ATRACT reconstruction (shown by the arrows in Fig. 6(B) and Fig. 7(B)) while not observed in 3D ATRACT. The profiles along the yellow-dashed line shown in Fig. 8 also demonstrate this observation.

V. Conclusion

In this paper, we presented a novel method that adapts the previously suggested ATRACT method in three dimensional by decomposing the standard ramp filter into the 3D Laplace filter and a 3D convolution-based filter. As opposed to the native ATRACT, the new method is able to handle the off-center ROI reconstruction caused by an asymmetric collimation. However, the 3D convolution is more computationally demanding than its 2D counterpart, which would consequently affect the reconstruction speed.

Acknowledgment

The authors gratefully acknowledge funding by Siemens AG, Healthcare Sector and of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative. The concepts and information in this paper were never submitted, published, or presented before.

References

[1] F. Noo, R. Clackdoyle, and J. D. Pack, "A Two-step Hilbert Transform Method for 2D Image Reconstruction," *Physics in Medicine and Biology*, vol. 49, no. 17, pp. 3903–3923, 2004.

[2] X. Pan, Y. Zhou, and D. Xia, "Image reconstruction in peripheral and central region-of-interest and data redundancy," *Medical Physics*, vol. 32, no. 3, pp. 673–684, 2005.

[3] H. Kudo, M. Courdurier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Physics in Medicine and Biology*, vol. 52, no. 9, p. 2207, 2008.

[4] B. Ohnesorge, T. Flohr, K. Schwarz, J. P. Heiken, and J. P. Bae, "Efficient correction for CT image artifacts caused by objects extending outside the scan field of view," *Medical Physics*, vol. 27, no. 1, pp. 39–46, 2000.

[5] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. J. Myers, "A novel reconstruction algorithm to extend the CT scan field-of-view," *Medical Physics*, vol. 31, no. 9, pp. 2385–2391, 2004.

[6] J. D. Maltz, S. Bose, H. P. Shukla, and A. R. Bani-Hashemi, "CT truncation artifact removal using water-equivalent thicknesses derived from truncated projection data," in *IEEE Engineering in Medicine and Biology Society*, pp. 2905–2911, 2007.

[7] A. Maier, B. Scholz, and F. Dennerlein, "Optimization-based Extrapolation for Truncation Correction," in *2nd CT Meeting*, pp. 390–394, 2012.

[8] F. Dennerlein, "Cone-beam ROI reconstruction using the Laplace operator," in *Proceedings of Fully 3D 2011*, pp. 80–83, 2011.

[9] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone beam algorithm," *Optical Society of America*, vol. 1, pp. 612–619, 1984.

[10] F. Dennerlein and A. Maier, "Region-of-interest reconstruction on medical C-arms with the ATRACT algorithm," in *Proceedings of SPIE*, p. 83131B, 2012.

[11] Y. Xia, A. Maier, F. Dennerlein, H. G. Hofmann, and J. Hornegger, "Efficient 2D filtering for cone-beam VOI reconstruction," in *IEEE NSS-MIC*, pp. 2415–2420, 2012.

Application of Image-Based Registration Method for Simultaneous Compensation of Cardiac and Respiratory Motions in Dual Gated Myocardial Perfusion SPECT

Taek-Soo Lee, Tao Feng, and Benjamin M. W. Tsui

***Abstract*—We developed a 4D SPECT image reconstruction method with simultaneous cardiac and respiratory (R&C) motion compensation based on the image-based registration method for improved 4D gated MP SPECT images and evaluated its performance using realistic simulated R&C dual gated myocardial perfusion (MP) SPECT projections. Using the 4D XCAT (eXtended CArdiac Torso) phantoms, a cycle of respiratory motion (RM) was simulated as 24 equally-spaced time frames while a cycle of cardiac beating motion was divided into 48 equally-spaced time frames for each of the 24 respiratory phases to simulate simultaneous R&C dual gating scheme. Almost noise-free projection data were generated using the SimSET simulation techniques that include the effect of collimator detector response, photon attenuation and scatter to simulate a typical 99mTc Sestamibi MP SPECT projection dataset, and were scaled and combined to form 6 equal amplitude respiratory gates and 8 equal time cardiac gates for R&C motions, respectively. Poisson noise was then added to model clinically acquired data. The noisy projection dataset was reconstructed using the 3D OS-EM with attenuation correction. Using a group-wise B-spline non-rigid image-based registration method with the iterations of the gradient-descent line search, the deformation fields of the R&C motions were estimated and the estimated deformation fields of the respiratory motion (RM) were applied to each cardiac phase of the dual gated SPECT images for the RM correction. Then, the RM compensated gated MP SPECT images were transformed by using the estimated deformation fields of the cardiac motion. Total Error was calculated to evaluate how much the reconstructed image is deviated from its phantom. The results showed that while the RM compensation improved the dual gated SPECT image quality in terms of reducing the motion blurring compared with the 3D OS-EM with no compensation, additional cardiac motion compensation with the optimized multi-resolution spacing of the control points further reduced the noise level significantly. We conclude that the proposed 4D SPECT image reconstruction method with R&C motion estimation and compensation allowed significant reduction of image blurring and noise level due to the dual R&C motions in the dual gated SPECT images. The realistically simulated dual R&C gated MP SPECT projection data set provides a powerful tool for developing a 4D image reconstruction and correction methods and the evaluation of their effects for clinical R&C gated MP SPECT/CT study.**

Index Terms— respiratory and cardiac motion, dual gating, compensation, 4D image reconstruction, SPECT

I. Introduction

It is well known that cardiovascular and respiratory motions are a major source of image artifacts in myocardial imaging. Thus, the importance of corrective 4D image reconstruction method with motion compensation has been emphasized in emission computed tomography.

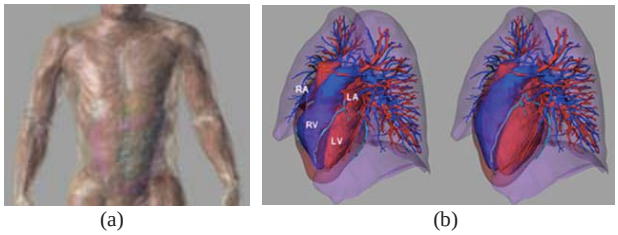
Recently, there have been clinical attempts for investigating the feasibility of dual cardiac-respiratory gating in emission computed tomography acquisition (ECT). However, such a task is extremely complex due to the many parameters that it entails.

In this study, we develop a realistic 4D simulation dataset that can model various cardiac and respiratory (R&C) gating schemes to study the effects of the motions, and also develop a 4D SPECT image reconstruction method with R&C motion compensation and evaluate its performance using the R&C dual gated dataset for improved accuracy and precision in 4D gated myocardial perfusion SPECT images.

II. Methods

A. The 4D Dual-Gating Dataset

The four-dimensional (4D) eXtended CArdiac-Torso (XCAT) phantom was used to simulate simultaneous dual gated R&C motions (Fig. 1). The respiratory motion (RM) excluding the heart was modeled by a total of 24 equally-spaced time frames of phantoms over a respiratory cycle (with a period of 5 seconds). The body phantom includes six separate organs; lung, liver, stomach, kidney, gall bladder, and background body.



(a) (b)

Fig. 1. XCAT phantom (a) anterior view of the XCAT torso, (b) heart phantoms with end-systolic (left) and end-diastolic (right) phases.

This work was supported by NIH grants R01 EB 000168 and R01 HL 068075.

Taek-Soo Lee, Ph.D. (Telephone: (443) 287-2006, E-mail: tslee@jhmi.edu), Tao Feng (Telephone: (443) 287-7314, E-mail: tfeng@jhu.edu), and Benjamin M. W. Tsui, Ph.D. (Telephone: (443) 287-4025, E-mail: btsui1@jhmi.edu) are with the Department of Radiology, Johns Hopkins University, Baltimore, MD 21287-0859, USA.

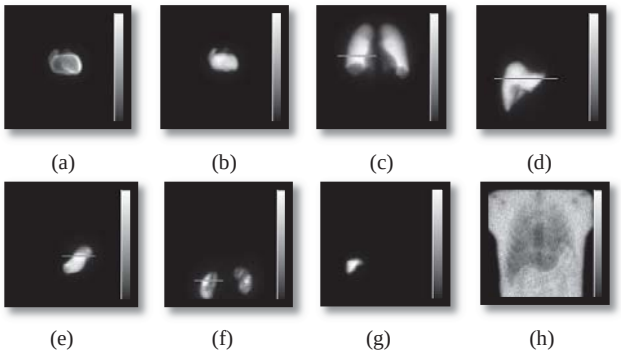
The beating heart which is consisted of myocardium and bloodpool was simulated separately with 48 equally-spaced time frames per cycle (with a period of 1 second) for each of 24 respiratory phases so that we can generate any combination of cardiac-respiratory gating schemes. Table 1 shows summary of the dataset.

TABLE I. THE 4D DUAL-GATING DATASET OF THE 4D XCAT PHANTOM

	Cardiac Dataset	Respiratory Dataset
Phantom	Heart only (myocardium, blood pool)	Lung, liver, stomach, kidney, gall bladder, background body
Frames/cycle	48/1 second	24/5 seconds
Possible Gating Schemes	8, 16, 24 equally space time gates/cycle	3, 4, 6 equally space time gates/cycle

B. Generation of Projection Data

As shown in Fig. 2, an almost noise-free projection dataset modeling a typical Tc-99m Sestamibi SPECT study was generated for the entire 4D XCAT phantom which includes the heart, blood pool, lungs, liver, kidneys, stomach, gall bladder, and remaining body using a long Monte Carlo simulation that include the effect of collimator detector response, photon attenuation and scatter. The separate projection datasets for body organs and the heart were combined to represent the complete torso of the body.



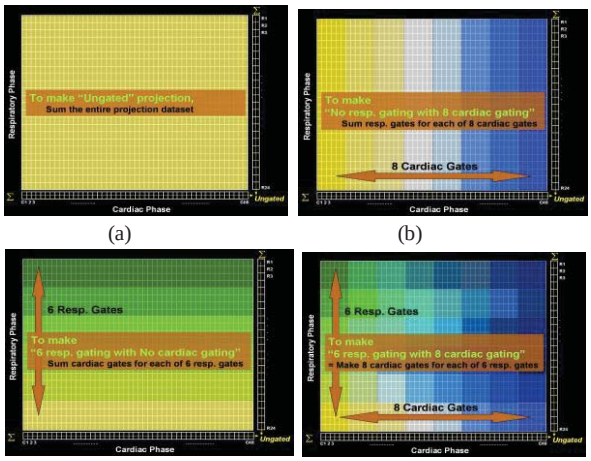
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2. Anterior view of the projection data generated using Monte Carlo simulation for different organs (a) myocardium, (b) blood pool, (c) lung, (d) liver, (e) stomach, (f) kidney, (g) gall bladder, (h) background body.

The entire projection dataset is able to effectively model the simultaneous dual R&C gated ECT/CT. As shown in Fig. 3, each grid of the 24 x 48 matrix represents a frame of a specific phase of R&C cycle. Using the master dataset, various kinds of R&C gating schemes can be generated by different combinations of R&C phases, and some sample gating schemes are shown in Fig.3. In this study, 6 equal amplitude gating and 8 equal time gating were used for the respiratory and cardiac motions, respectively.

C. The group-wise B-spline non-rigid image-based registration method



(a) (b)

(c) (d)

Fig. 3. Various respiratory-cardiac gating schemes which can be generated from the dataset. Each different color block represents one frame. (a) ungated, (b) No respiratory gating with 8 frames cardiac gating, (c) 6 frames respiratory gating with no cardiac gating, (d) simultaneous 6 frames respiratory and 8 frames cardiac gatings.

The group-wise B-spline non-rigid image-based registration method was adopted for RM estimation based on respiratory-gated images. The assumption for the method is that there exists an underlying reference frame, which is the "middle point" of all the images to be registered. It is formulated as the following:

$$\left\{\hat{\vec{T}}_g(\vec{r}),\, g=1,2,...,G\right\} = \arg\min \sum_{g_2\neq g_1}^{G}\sum_{g_1=1}^{G}\iiint_{\Omega}\left[f_{g_1}\left(\vec{T}_{g_1}(\vec{r})\right) - f_{g_2}\left(\vec{T}_{g_1}(\vec{r})\right)\right]^2 dxdydz \quad (1)$$

where $\vec{T}_g(\vec{r})$ is a transform function modeling the respiratory motion from the reference frame to frame g, and $f_g(\vec{r}_g)$, $g=1,2,...,G$ are the continuous representations for the respiratory-gated images of G frames, and $\vec{r}=(x,y,z)$ is the spatial coordinate for the reference frame defined by the following constraint: $\vec{r}=\frac{1}{G}\sum_{g=1}^{G}\vec{T}_g(\vec{r})$. The spatial position of every point in the reference frame is the mean or average of its corresponding positions in all frames or images. If the spatial positions of a point in all frames distribute along a straight line, then its position in the reference frame is the middle point of the line. Here, we modeled $\vec{T}_g(\vec{r})$ as the cubic B-spline non-rigid transform which has been widely used in the non-rigid image registration of medical images from the same or different modalities [6, 7].

$$\vec{T}_g(\vec{r}) = \vec{r} + \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)(\alpha^g_{x,i+l}, \alpha^g_{y,j+m}, \alpha^g_{z,k+n}) \quad (2)$$

where $(\alpha^g_{x,i+l}, \alpha^g_{y,j+m}, \alpha^g_{z,k+n})$ are the displacements of the control points to be estimated in the x, y and z directions, respectively for frame *g*, and $B_l(u)$ is the l^{th} basis function of the B-spline. Other notations in (2) are as the following:

$$i = [x/d_x]\text{-}1,\ j = [y/d_y]\text{-}1,\ k = [z/d_z]\text{-}1 \quad (3)$$

$$u = x/d_x\text{-}[x/d_x],\ v = y/d_y\text{-}[y/d_y],\ w = z/d_z\text{-}[z/d_z] \quad (4)$$

where $[x]$ is the largest integer less than the real number x and (d_x,d_y,d_z) is the initial spacing of the control points. $B_l(u)$ is the l^{th} basis function of the B-spline. The B-spline basis functions of order 0 to 3 are defined as the following:

$$\begin{aligned} B_0(u) &= (1\text{-}u)^3/6,\quad B_1(u) = \left(3u^3\text{-}6u^2+4\right)/6 \\ B_2(u) &= \left(\text{-}3u^3+3u^2+3u+1\right)/6,\quad B_3(u) = u^3/6 \end{aligned} \quad (5)$$

The gradient-descent line-search optimizer was employed to search for the optimal displacements of control points which minimized the cost function in (1) under its constraint. The gradient of the cost function was analytically computed and a line search was performed along the gradient direction to determine the step size. The iterative minimization process was terminated after a pre-set number of iterations.

To reduce the computational cost, we employed a multi-resolution scheme for the spacing of the B-spline control points. In the first stage of registration, we started with 4 x 4 x 4 control points in the central regions of the extracted image from the entire SPECT images. Then we increased the number of control points to 6 x 6 x 6 for the second stage and further to 8 x 8 x 8 for the third stage. At each stage of the control-point spacing, we sampled the images also in a multi-resolution scheme using 'level' for the cost function evaluation and gradient computation to further reduce the computation time. As we moved up to a higher level at a given stage, the number of B-spline control points was doubled. For example, starting from 4 x 4 x 4 level 1 and 2 would increase B-spline control points to 8 x 8 x 8, and from 8 x 8 x 8 and 16 x 16 x 16, respectively.

The noisy projection dataset was reconstructed using the 3D OS-EM with attenuation correction with an attenuation map averaged over the RM. The deformation fields of the R&C motions were estimated and the estimated deformation fields of the RM were applied to each cardiac phase of the dual R&C gated SPECT images for the RM correction. Finally, the RM compensated gated MP SPECT images were transformed by using the estimated deformation fields of the cardiac motion.

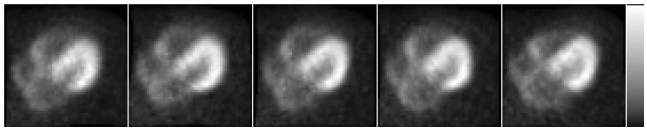
The total error for the *i*th cardiac gated image was calculated for the region-of-interest (ROI), a cubic volume which encompassed the entire heart (ROI_{heart}), in the reconstructed image compared to its corresponding phantom frames (true). During the calculation of the total error, the reconstructed and phantom images were normalized individually by dividing by its mean so that the differently scaled images could be compared. The total error for the *i*th gate is defined as following;

$$\text{Total Error}_i = \frac{1}{n}\sum_{j\in ROI_{heart}}\left(\frac{x_j^i}{\bar{x}_j} - \frac{\lambda_j^i}{\bar{\lambda}_j}\right)^2 \quad (6)$$

where x_j^i is *j*th pixel value of the *i*th cardiac gate of the reconstructed image, λ_j^i is *j*th pixel value of the *i*th cardiac gate of the phantom slice, $\bar{x}_j$ is the mean of ROI_{heart} of the *i*th cardiac gate of the reconstructed image, $\bar{\lambda}_j$ is the mean of ROI_{heart} of the *i*th cardiac gate of phantom slice, and *n* is the number of pixels in the ROI_{heart}.

III. RESULTS

We first optimized the multi-resolution spacing of the control points in the group-wise B-spline non-rigid image registration method. Fig. 4 shows reconstructed images of the heart in the end-systolic phase with R&C motion compensated. It demonstrates that the resolution of the B-spline control-points spacing in terms of the number of grid and level affect the accuracy of the motion estimation.

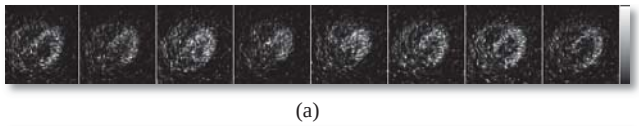


(a) (b) (c) (d) (e)

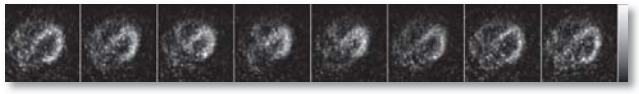
Fig. 4. Reconstructed images of a cardiac frame in the end-systolic phase with R&C motion compensated. (a) Grid 4 with 2 levels, (b) Grid 6 with 1 level, (c) Grid 6 with 2 levels, (d) Grid 8 with no level, and (e) Grid 8 with 1 level.

Based on the results, we employed 6 x 6 x 6 uniformly spaced B-spline control-points with 1 level and performed the gradient-descent line search for image-based RM estimation with the group-wise B-spline non-rigid image registration method.

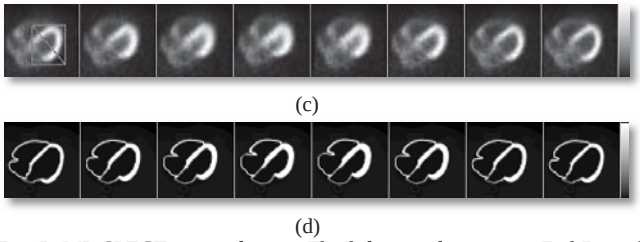
The RM compensation improved the dual gated SPECT image quality in terms of reducing the motion blurring compared with the 3D OS-EM with no compensation, as shown in Fig. 5 (a) and (b). Fig. 5 (c) demonstrates an additional cardiac motion (compensation further reduced the noise level drastically without compromising the motion blur.



(a)



(b)



(c)

(d)

Fig. 5. MP SPECT image frames. The left-most frame is in End-Diastolic phase and 4th frame from the left is in the End-Systolic phase. (a) No R&C motion compensation, (b) RM compensated only, (c) R & C motion compensated, and (d) corresponding image slices of the phantom.

Fig. 6 shows the profiles of the MP SPECT image frames in the Fig. 5(c). It demonstrated the movement of myocardium across the cardiac frames. However, since the R&C motion compensation technique was applied, the position of the heart was not significantly shifted through the cardiac cycle.

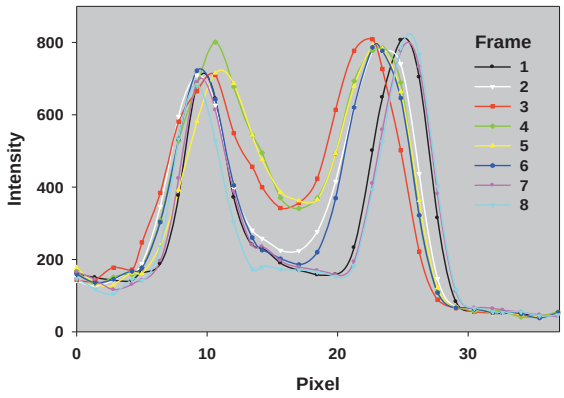


Fig. 6. Profiles of the MP SPECT image frames shown in the Fig. 5 (c) R&C motion compensated images.

Fig. 7 shows the significant reduction of the total error in the R&C motion compensated images as well as in respiratory motion only compensated images, compared to the no motion compensated images.

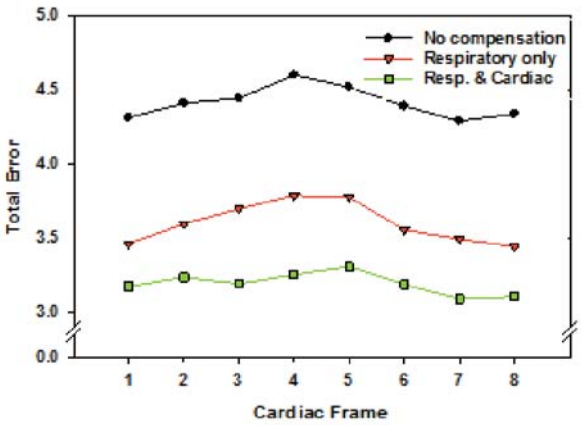


Fig. 7. Total error in the 4D reconstructed images with no motion compensation applied, respiratory motion compensation only applied, and R&C motion compensation applied.

IV. Conclusion

The proposed 4D SPECT image reconstruction method with R&C motion estimation and compensation with optimized parameters allowed significant reduction of image blurring and noise level due to the dual motions and dual gating in the cardiac gated SPECT images.

The realistically simulated simultaneous dual R&C gated MP SPECT projection data provides a powerful tool for developing image reconstruction and correction methods and the evaluation of their effects for clinical SPECT/CT dual gating study.

V. Acknowledgment

This work was supported by NIH grants R01 EB 000168 and R01 HL 068075. The content of this work is solely the responsibility of the authors and does not necessarily represent the official view of the PHS or its various institutes.

VI. References

[1] M. M. Ter-Pogossian, S. R. Bergmann, B. E. Sobel, "Influence of cardiac and respiratory motion on tomographic reconstructions of the hearts" implications for quantitative nuclear cardiology," *J Comput Assist Tomogr*, vol. 6(6), pp. 1148-1155, 1982.

[2] G. J. Klein, B. W. Reutter, M. H. Ho, J. H. Reed, R. H. Huesman, "Real-time system for respiratory-cardiac gating in positron tomography," *IEEE Transactions on Nucl Sci*, vol. 45(4), pp. 2139-2143, 1998.

[3] K. Cho, S. Kumiata, S. Okada. et al, "Development of respiratory gated myocardial SPECT system," *J Nucl Cardiol*, vol. 6, pp. 20-28, 1999.

[4] T.-S. Lee, W. Paul Segars, B. Tsui, "The Development and Application of a Realistic Simulation Dataset for Simultaneous Cardiac and Respiratory Gated ECT/CT," in *Nuclear Science Symposium and Medical Imaging Conference Record*, Honolulu, Hawaii, pp. 3656, 2007.

[5] K. Gould et al, "Frequent diagnostic errors in cardiac PET/CT due to misregistration of CT attenuation and emission PET images: a definitive analysis of causes, consequences, and corrections. *J Nucl Med*, vol. 48, pp.1112-1121, 2007.

[6] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Transactions on Medical Imaging, vol. 18, pp. 712-721, AUG 1999.

[7] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," IEEE Transactions on Medical Imaging, vol. 22, pp. 120-128, Jan 2003.

Optimizing image reconstruction for simultaneous sub-mm clustered pinhole PET-SPECT

Marlies C. Goorden, Frans van der Have, and Freek J. Beekman

***Abstract*— A newly developed Versatile Emission Computed Tomography system (VECTor) enables simultaneous imaging of SPECT and PET tracer molecule distributions at sub-mm resolutions in mice. VECTor uses a dedicated collimator with clusters of small opening-angle pinholes that is mounted on a SPECT system with stationary NaI detectors. The novel pinhole geometry and the extended energy range of imaged gamma photons require a new evaluation of image reconstruction software instead of only slightly adapting standard SPECT methods. The preliminary results presented in this paper demonstrate that such a reconstruction optimization strongly improves VECTor's performance.**

Projections from different pinholes slightly overlap on the 3 gamma detectors of VECTor. Near the edges of the pinhole projections, a small amount of mismodeling -which is often unavoidable- leads to a large loss of information about the emission direction of the detected gamma photons. We tested if uniformity of SPECT images was improved by simply ignoring detector pixels located near the pinhole projection edges. Secondly, we investigated if accurately modeling the varying depth-of-interaction (DOI) in the NaI detector and including a larger portion of the point spread function (PSF) tails improved PET resolution phantom images.

Reconstructed SPECT images of a syringe were most uniform when the detector pixels within a 1 to 2 pixel distance from the edges of the pinhole projections were ignored. Furthermore, we found a remarkable improvement in PET resolution phantom reconstructions when a large portion of the PSF tails was used in the reconstruction and when DOI was modeled. To conclude, we have shown that the performance of VECTor can be significantly improved by optimizing its image reconstruction software.

***Index Terms*—SPECT, PET, pinholes, fully 3D iterative reconstruction**

I. Introduction

SPECT and PET imaging of small animals has recently been combined in a single Versatile Emission Computed Tomography system (VECTor) [1]. VECTor attains cutting edge image resolution (<0.5 mm) for SPECT while simultaneously imaging PET tracers at sub-mm resolutions. An example of quadruple PET-SPECT tracer imaging in a mouse is displayed in Fig. 1. Since VECTor fully integrates PET and SPECT, the tracer distributions in such an image are perfectly aligned in space and time and can therefore be directly correlated. As a consequence, VECTor enables a whole new range of applications in biomedical research, notably applications in which the correlation of multiple biological processes is of importance.

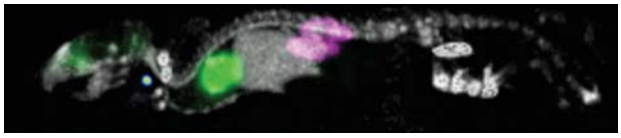


Fig. 1. Simultaneous PET-SPECT quadruple isotope imaging with VECTor. Maximum Intensity Projection are shown of a 60 minute total body mouse scan with 100 MBq ^{99m}Tc-HDP (Gray), 35 MBq ^{18}F-FDG (Green), 19 MBq ^{111}In-Pentetreotide (Magenta), and 5 MBq ^{123}I-NaI (Rainbow).

VECTor is based on a novel collimation technique that exploits clusters of pinholes [2, 3] with each cluster sampling the same field-of-view as a single traditional pinhole (Fig. 2(a)). Since each of the pinholes in the cluster has a relatively small opening angle, pinhole edge penetration –which would be a major problem when imaging 511 keV annihilation photons with a standard SPECT collimator- is reduced. The clustered pinhole collimator is placed into a SPECT system with stationary NaI gamma detectors (U-SPECT-II, MILabs BV, The Netherlands [4]).

A key ingredient to obtaining high-resolution images with complex pinhole geometries is the use of iterative reconstruction algorithms that (partly) compensate for effects like spatially variant sensitivity and resolution. Ideally, these algorithms are based on an accurate knowledge of point spread functions (PSFs), the system's response to a point source. For practical reasons, we require that VECTor is calibrated with a limited number of point source measurement of a single isotope (^{99m}Tc) but we model the energy dependent PSFs for each isotope individually.

Being a new modality, optimizing VECTor's image reconstruction software may have a significant impact on its performance. This is confirmed by the preliminary results that we present in this paper. We have focused on aspects in which VECTor differs from the standard U-SPECT-II collimators. First, in contrast to these standard collimators, the projections of the clustered pinholes onto the 3 gamma

M. C. Goorden is with Delft University of Technology, Section Radiation Detection and Matter, Mekelweg 15, 2629 JB Delft, The Netherlands (e-mail: m.c.goorden@tudelft.nl).

F. van der Have and F.J. Beekman are with Delft University of Technology, Section Radiation Detection and Matter, Mekelweg 15, 2629 JB Delft, The Netherlands and MILabs, Utrecht, The Netherlands.

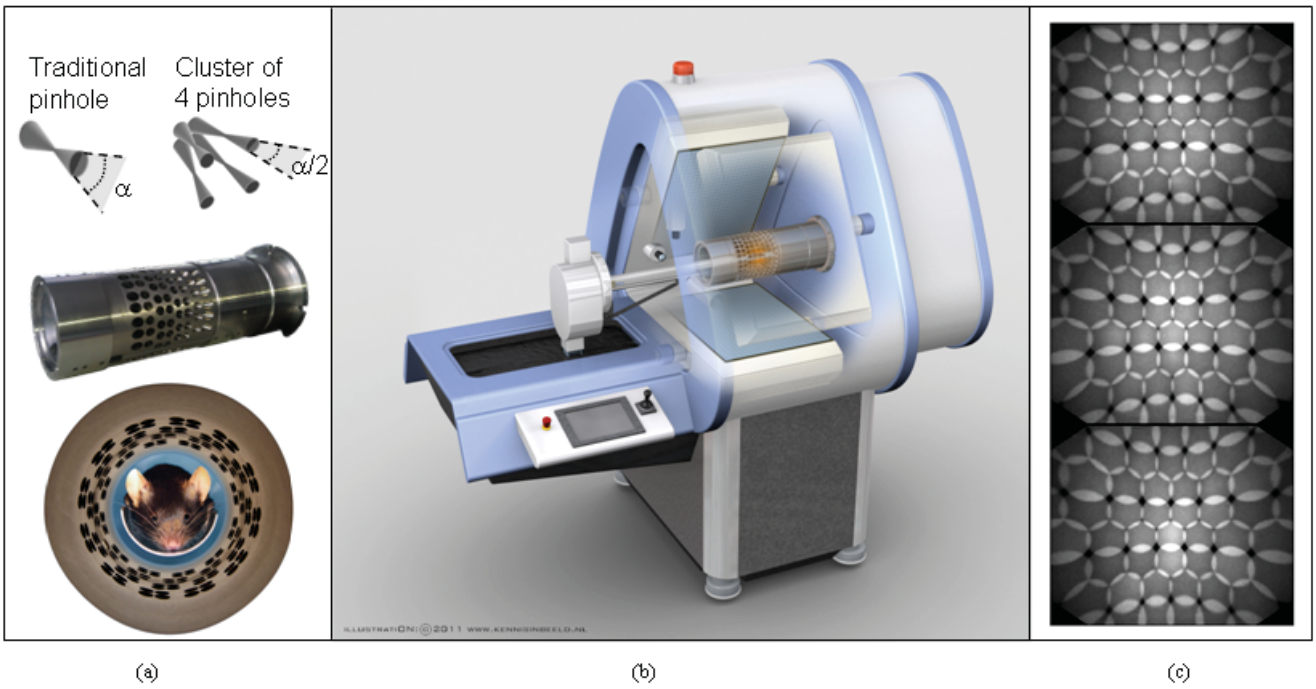


Fig. 2. Integration of the clustered multi-pinhole collimator into an existing SPECT/CT platform. (a) A traditional pinhole with opening angle α and a cluster of four pinholes with approximately the same field-of-view and opening angle α/2. The clustered multi-pinhole collimator optimized for imaging SPECT and PET tracers, into which a mouse is placed, contains 48 cluster of 4 pinholes each. (b) The collimator is mounted in a SPECT/CT system with three large-area gamma detectors. (c) Projection image of an extended ^{99m}Tc source onto the detector; pinhole projections slightly overlap. Subfigures (a) and (b) were reprinted from [1].

detectors of the system slightly overlap when an extended source is imaged (Fig. 2(c)). Note that across the edge of the ellipsoidal projection area belonging to a certain pinhole, a very abrupt change occurs in the probability that a detected gamma photon originated from that pinhole. This means that even a slight amount of mismodeling of the exact location of these edges, which can occur due to manufacturing uncertainties, leads to a considerable loss of information about the emission direction of gamma photons detected near these edges. Therefore, for some detector pixels in these edge regions, the sensitivity that is gained when their signal is used in the image forming process may be completely offset by the ambiguous information they carry. With this in mind, we investigated if it helped to simply ignore the signal of certain detector pixels in image reconstruction. Secondly, high-energy annihilation gammas have broad PSFs with longer tails than is common in SPECT as well as an increased variation of the depth-of-interaction (DOI) in the gamma detector. We investigated whether including the tails and the varying DOI into the PSF model visually improved PET images.

II. METHODS

A. Pinhole Geometry and System description

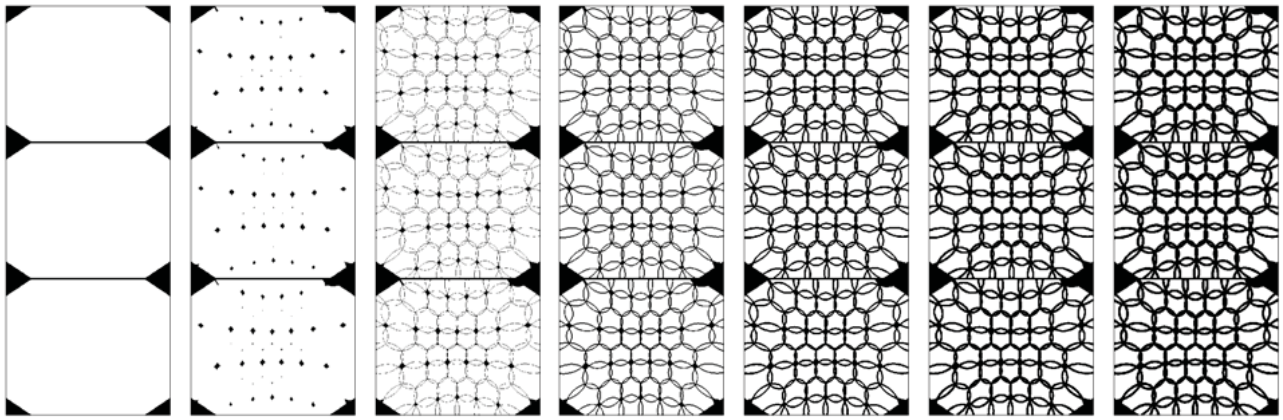
The clustered pinhole collimator (Fig. 2(a)) contains 48 clusters of four pinholes placed in four rings. It was integrated into the U-SPECT-II/CT system [4] which has three large-area NaI(Tl) gamma detectors placed in a triangular set-up (Fig. 2(b)). The projection of an extended ^{99m}Tc source on the 3 gamma detectors is shown in Fig. 2(c); projections of different pinholes slightly overlap. A detailed description of the collimator geometry can be found in [1].

B. Image reconstruction and calibration

The scanning focus method [5] was used for data acquisition. A Pixel-based Ordered Subsets Expectation Maximization (POSEM) algorithm with 32 subsets was used for image reconstruction [6]. Scatter was corrected for by using a standard triple energy window correction [7].

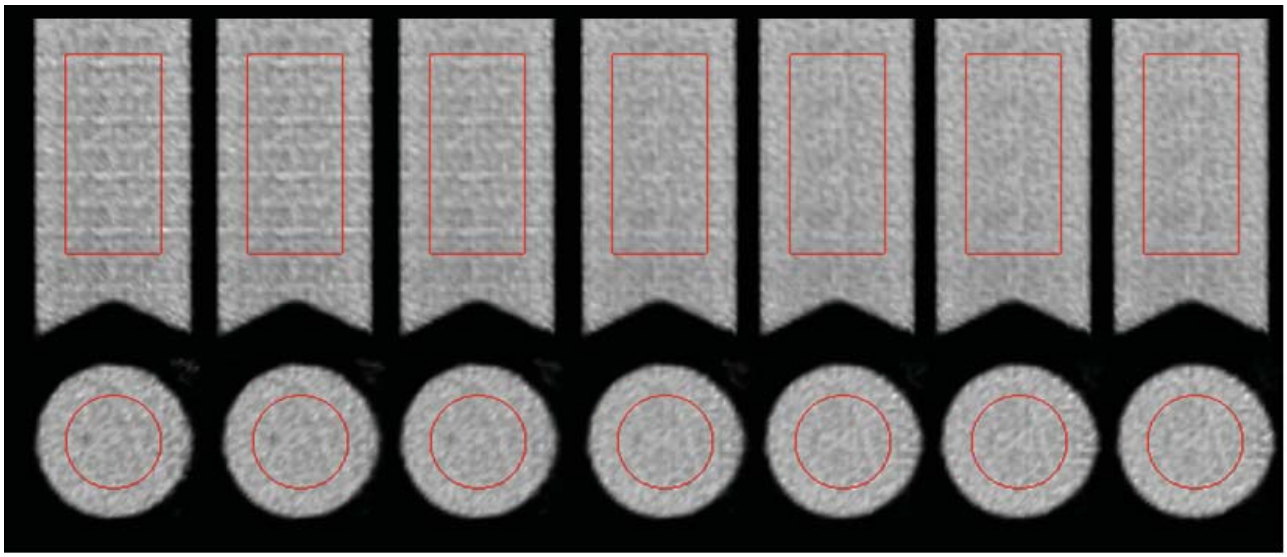
A ^{99m}Tc point source measurement [8] was done to correct for small geometrical misalignments that can occur due to slight rotations and translations of the collimator or the gamma detectors with respect to the expected (designed) position and orientation. The detector position model used in the fit was based on projecting a point source location through the pinhole centers on the detector plane [8]. Since an arbitrary translation + rotation can be described by 6 parameters, we fitted 24 system parameters to the point source experiment (3 detectors + 1 collimator). It was assumed that relative pinhole positions and orientations in the collimator were as designed.

Based on this fit, PSFs were calculated with a ray tracing code that calculated the path length L of gamma photons through the collimator material. In our code, we could set a cut-off C which had the effect to only save those parts of the PSFs with attenuation $\exp(-\mu L)>C$. Here μ is the collimator's energy dependent attenuation coefficient. Furthermore, the varying DOI in the gamma detector was modeled by also using a ray tracing code for the NaI scintillator. In our code, DOI modeling could be turned on/



(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Different detector masks that were applied in system matrix generation; the black areas represent those pixels that were *not* used in image reconstruction. (a) The active area of the 3 gamma detectors was fully used. (b) Pixels not contained in any pinhole projection were ignored. (c) Pixels at the edges of the pinhole projections were also left out. In subsequent masks, pixels within a 1, 2, 3 and 4 pixel distance were also ignored (d-g).



Uniformity: 17.5% (a) 15.4% (b) 13.7% (c) 12.9% (d) 13.0% (e) 13.2% (f) 13.5% (g)

Fig. 4. Two mutually perpendicular slices (slice thickness 2mm) through reconstructions of a 12 ml syringe filled with ^{99m}Tc. Reconstructions (a)-(g) correspond to detector masks of Fig. 3 with the same label. Uniformity in the cylindrical region-of-interest (red line) is provided below each reconstruction as a percentage of the mean activity in the same region.

off at will. When DOI modeling was used in the ray tracing code, a DOI correction was also used in the detector position model for the point source fit. Detector resolution was modeled by a Gaussian detector response with 3.5 mm FWHM. Finally, our code read in a detector mask in which some of the detector pixels could be set to zero and therefore not be incorporated in the system matrix.

C. Reconstruction optimization

The pinhole projection areas for each pinhole were determined by means of fitting overlapping ellipses to the measured ^{99m}Tc flood source projection (Fig. 2(c)), where the position of the center, length of major and minor axes and angle of each of the ellipses were fit parameters. We tested different detector masks (Fig. 3). The basis was formed by mask (a) which included all active detector pixels. We subsequently left out more and more pixels; first those that were not located in at least 1 of the ellipsoidal pinhole projections (b), subsequently pixels at the edges of the pinhole projections (c) and then all pixels within a 1, 2, 3 or 4 pixel distance from the edges (d-g). Compared to Fig. 3(a), masks (b-g) contained 97%, 92%, 82%, 73%, 67%, 62% of the pixels, respectively. We tested how use of these masks influenced uniformity in reconstructions of a 12 ml syringe filled with 225 MBq ^{99m}Tc and scanned for 9 hours. These tests were done for a cut-off C=20% and no DOI modeling.

Secondly, we lowered C from 20% to 1% and subsequently incorporated DOI modeling. For this, we assumed an attenuation coefficient of 0.012 mm^{-1} in the NaI scintillator based on previous simulations and experiments [9]. We tested how changing these parameters influenced PET resolution phantom images. To this purpose, a Jaszczak resolution phantom with capillary diameters of 0.7-1.5 mm was filled with 102 MBq ^{18}F solution and

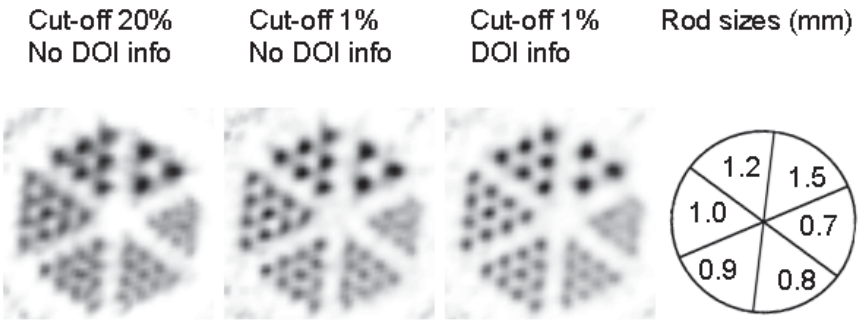


Fig. 5. PET images of a resolution phantom. (a) System matrix generated with a cut-off of 20% and no DOI modeling. (b) System matrix generated with a lower cut-off of 1%, DOI modeling not included. (c). Cut-off kept at 1%, DOI modeling included. (d) Rod sizes in each segment. Slice thickness of the reconstructions was 2 mm.

scanned for 4 hours. The distance between capillary centers was equal to twice the capillary diameter.

RESULTS

In Fig. 4 we show slices through the ^{99m}Tc uniform phantom reconstructions for all tested detector masks. Below each of the reconstructions, uniformity in a cylindrical region-of-interest (also shown) is provided as a percentage of the mean activity in the same region. Clearly, line-shaped artifacts are present in Fig. 4(a) which uses all detector pixels in the active detector area. As one does not use detector pixels in areas outside the pinhole projections (Fig. 4(b)) and leaves out more and more pixels near the edges (Fig. 4(c)-4(g)), these artifacts gradually disappear. Both visual inspection and calculated uniformity indicate that there is an optimum; it is best to leave out all pixels within a 1 or 2 pixel distance from the edges (Fig. 4(d) and 4(e)).

PET images of the resolution phantom are shown in Fig. 5. Lowering the cut-off from 20% (a) to 1% (b) resulted in images which appear less noisy. However, the rods in 5(a) and 5(b) are somewhat triangularly shaped. Including DOI modeling into the matrix generation obliterates this effect (c). In the latter figure there is no visual deformation of rods.

III. DISCUSSION AND CONCLUSION

In this paper, we have very convincingly demonstrated that optimizing image reconstruction software of VECTor can significantly improve image quality. We found that uniformity in SPECT images improves if one ignores the gamma photons detected in detector pixels near the edges of the pinhole projections. Images with the best uniformity only used 70 to 80% of the detector pixels. Apparently, the increased sensitivity that is obtained if these gamma photons would be taken into account does not outweigh the ambiguous information they carry. This is reminiscent of what other studies have found on the effect of overlapping pinhole projections on reconstructions [10, 11].

Furthermore, we found that for imaging high-energy annihilation photons resulting from PET tracers, it is important to model large portions of the PSF tails and to also model the variable DOI in the gamma detector. Such modeling has a significant impact on the appearance and visibility of small structures.

Here we have presented preliminary results. In the near future we plan to extend our study to systematically investigate the influence and interplay of these parameters on SPECT, PET and combined SPECT-PET images of several phantom and animal scans.

REFERENCES

[1] M. C. Goorden, F. van der Have, R. Kreuger, R. M. Ramakers, B. Vastenhouw, J. P. H. Burbach, J. Booij, C. F. M. Molthoff, and F. J. Beekman, "VECTor: A Preclinical Imaging System for Simultaneous Submillimeter SPECT and PET," Journal of Nuclear Medicine, vol. 54, no. 2, pp. 306-312, 2013.

[2] F. J. Beekman, "A focused pinhole gamma detection device," EP20730392009

[3] M. C. Goorden and F. J. Beekman, "High-resolution tomography of positron emitters with clustered pinhole SPECT," Physics in Medicine and Biology, vol. 55, no. 5, pp. 1265-1277, 2010.

[4] F. van der Have, B. Vastenhouw, R. M. Ramakers, W. Branderhorst, J. O. Krah, C. Ji, S. G. Staelens, and F. J. Beekman, "U-SPECT-II: an ultra-high-resolution device for molecular small-animal imaging," Journal of Nuclear Medicine, vol. 50, no. 4, pp. 599-605, 2009.

[5] B. Vastenhouw and F. J. Beekman, "Submillimeter total-body murine imaging with U-SPECT-I," Journal of Nuclear Medicine, vol. 48, no. 3, pp. 487-493, 2007.

[6] W. Branderhorst, B. Vastenhouw, and F. J. Beekman, "Pixel-based subsets for rapid multi-pinhole SPECT reconstruction," Physics in Medicine and Biology, vol. 55, no. 7, pp. 2023-2034, 2010.

[7] K. Ogawa, Y. Harata, T. Ichihara, A. Kubo, and S. Hashimoto, "A practical method for position-dependent Compton-scatter correction in single photon-emission ct," IEEE Transactions on Medical Imaging, vol. 10, no. 3, pp. 408-412, 1991.

[8] F. van der Have, B. Vastenhouw, M. Rentmeester, and F. J. Beekman "System calibration and statistical image reconstruction for ultra-high resolution stationary pinhole SPECT," IEEE Transactions on Medical Imaging, vol. 27, no. 7, pp. 960-971, 2008.

[9] M. C. Goorden, F. van der Have, R. Kreuger, and F. J. Beekman, "An efficient simulator for pinhole imaging of PET isotopes," Physics in Medicine and Biology, vol. 56, no. 6, pp. 1617-1634, 2011.

[10] K. Vunckx, P. Suetens, and J. Nuyts, "Effect of overlapping projections on reconstruction image quality in multipinhole SPECT," IEEE Transaction on Medical Imaging, vol. 27, no. 7, pp. 972-983, 2008.

[11] G. S. Mok, B. M. Tsui, and F. J. Beekman, "The effects of object activity distribution on multiplexing multi-pinhole SPECT," Physics in Medicine and Biology, vol. 56, no. 8 pp. 2635-2650, 2011.

Comparison of Gaussian and non-isotropic adaptive projection filtering for rotational 3D X-ray angiography

Dirk Schäfer, Peter van de Haar, Michael Grass

Abstract— **We present a noise-adaptive, edge-preserving projection filter with subsequent filtered back-projection (FBP) reconstruction for 3D rotational angiography. Contrast agent filled vessels are high contrast objects, which favor the application of edge preserving filtering in the projection domain. We propose a method based on minimization of the total variation (TV) and compare the performance to isotropic Gaussian filtering at equal background noise. For the phantom data the visibility of small vessels is better preserved with the TV filter compared to Gaussian filtering at equal background noise level. This behavior is confirmed on a clinical liver angiography data set.**

I. Introduction

Noise and noise streaks are a limiting factor of image quality in tomographic X-ray imaging especially in large and obese patients or for very low dose acquisitions. High noise levels in the X-ray projections translate to 3D streak artifacts and high 3D noise levels. Adaptive projection or sinogram filtering has been investigated already for a long time and is still of interest due to increasing dose awareness.

Hsieh derived the link of noise characteristics of projection data before and after logarithmic transform and developed an adaptive mean filter to be applied before log-transform [4] for noise reduction in CT imaging. Isotropic triangular filter kernels have been used in all three sinogram directions [5], or isotropic Gaussian kernels [11] have been applied in 2D projection space. Manduca et. al. investigated edge preserving bi-lateral filtering for CT dose reduction [6] and recently Maier et al. applied anisotropic adaptive 3D sinogram filtering to cone beam CT data [7]. They used the computationally complex and time-consuming estimation of local orientations with the structure tensor to steer their filter. Another computationally efficient method for denoising is the minimization of the total variation (TV) of an image [8], which has been used extensively in the last years for TV based volume regularization in iterative reconstruction (e.g. [9]). Projection or sinogram filtering with TV methods has been received much less attention. Zanella et. al. applied a TV based regularization method to dental radiographs [12]. Brown et. al. used TV projection filtering to obtain a better start image for iterative reconstruction for improved convergence [1]. They also showed that reconstructing with a combined projection and image TV approach is nearly equivalent to performing a full iterative reconstruction algorithm for a low dose simulation.

In this article, we present an adaptive noise removal algorithm based on edge preserving Huber regularized TV minimization (hTV) in the projection domain as pre-filtering step with subsequent FDK reconstruction [3] for rotational 3D X-ray angiography. The goal is to remove noise while preserving even small high contrast vessels. The hTV filter is compared to isotropic Gaussian filtering at equal background noise. In Sec. II the filtering methods are presented and evaluated in Sec. III on phantom and clinical data.

DS and MG are with Philips Research Europe - Hamburg, Röntgenstraße 24–26, 22335 Hamburg, Germany. PvdH is with Philips Healthcare, Veenpluis 6, 5684 PC Best, The Netherlands. corresponding author: dirk.schaefer@philips.com

II. Methods

A. Isotropic adaptive filtering

The raw projection intensity data is denoted $g(u,v)$, where u is the coordinate corresponding to the number of columns and v to the number of rows. The denoised data $g_{\mathrm{ag}}(u,v)$ is adaptively filtered with Gaussian kernels [11] with variance s:

$$g_{\mathrm{ag}}(u,v) = g_{\mathrm{ag}}(u,v:t,s) = g_{\mathrm{ag}}(t,s). \tag{1}$$

Following the reasoning in [4] only intensities below the threshold t are filtered. Because the after-log noise variance $\sigma^2_{\mathrm{after-log}}$ of the projection data is inversely proportional to the measured Poisson signal g (neglecting electronic noise) [4]

$$\sigma^2_{\mathrm{after-log}} = \frac{1}{g},$$

we choose the Gaussian parameter s inversely proportional to the signal $s \propto 1/g$.

B. TV based adaptive filtering

To find a denoised version z of the projection g, we minimize the total variation of the signal, while keeping the quadratic deviation to the original signal small:

$$\underset{z}{\operatorname{argmin}} \left[\int \Psi_\delta(\nabla z)\,\mathrm{d}u\mathrm{d}v \;+\; \lambda \int (z-g)^2\,\mathrm{d}u\mathrm{d}v \right]. \tag{2}$$

The Huber-function Ψ_δ is chosen as an approximation of the absolute value function to avoid discontinuities in the first derivatives [10]:

$$\Psi_\delta(\nu) = \begin{cases} \dfrac{1}{2}\nu^2 & \text{for } |\nu| \le \delta \\ \delta \times \left(|\nu| - \dfrac{\delta}{2} \right) & \text{otherwise} \end{cases}, \tag{3}$$

where δ is a positive parameter. The corresponding Euler-Lagrange partial differential equation is obtained by finding

the zeros of the partial derivatives of the cost function with respect to z:

$$0 = -\nabla \cdot \frac{\partial}{\partial z}\Psi_\delta(\nabla z) + \lambda(z-g) \text{ with } \left.\frac{\partial z}{\partial \mathbf{n}}\right|_{\partial\Omega}, \quad (4)$$

where $\mathbf{n}$ denotes the (vanishing) outer normal along the boundary $\partial\Omega$. Using the derivative of $\Psi(\nabla z)$

$$\frac{\partial}{\partial z}\Psi_\delta(\nabla z) = \begin{cases} \nabla z & \text{for} |z| \le \delta \\ \delta\dfrac{\nabla z}{|\nabla z|} & \text{otherwise} \end{cases}, \quad (5)$$

one gets:

$$0 = -\nabla \cdot \left[\delta \frac{\nabla z}{\max(|\nabla z|, \delta)}\right] + \lambda(z-g). \quad (6)$$

The update formula can be derived following the discretization steps in [2] for a pixel O and the corresponding neighborhood Λ_o with the directions E, N, W, S:

$$z_o^{(n+1)} = \frac{\delta \sum_{p\in\Lambda_o} w_p^{(n)} z_p^{(n)} + \lambda g_o}{\delta \sum_{q\in\Lambda_o} w_q^{(n)} + \lambda}, \quad (7)$$

with the weights:

$$w_p = \frac{1}{\max(|\nabla z_p|, \delta)} \quad p \in \Lambda_o = \{E, N, W, S\}. \quad (8)$$

Following the rationale in [1], [2] we set the parameter λ inversely proportional to the after-log noise variance and filtering is only applied to regions below a certain threshold t like the adaptive Gaussian filtering. The hTV-filtered projection g_{hTV} is then obtained after a pre-defined number of iterations I:

$$g_{\mathrm{hTV}}(u, v : t, I, \delta) = g_{\mathrm{hTV}}(t, I, \delta) = z_o^{(I)}, \quad (9)$$

with the update formula from eq. 7.

III. Results

The adaptive filtering methods presented in Sec. II are evaluated on a vessel phantom in a water bath and on abdominal angiographic patient data using a flat-detector C-arm system (Allura Xper FD20, Philips Healthcare, Best, The Netherlands). The filter settings have been adjusted in such a way that equal noise levels in homogeneous regions of the reconstructed volume are obtained. This allows to compare the visibility and sharpness of the vessels at equal background noise.

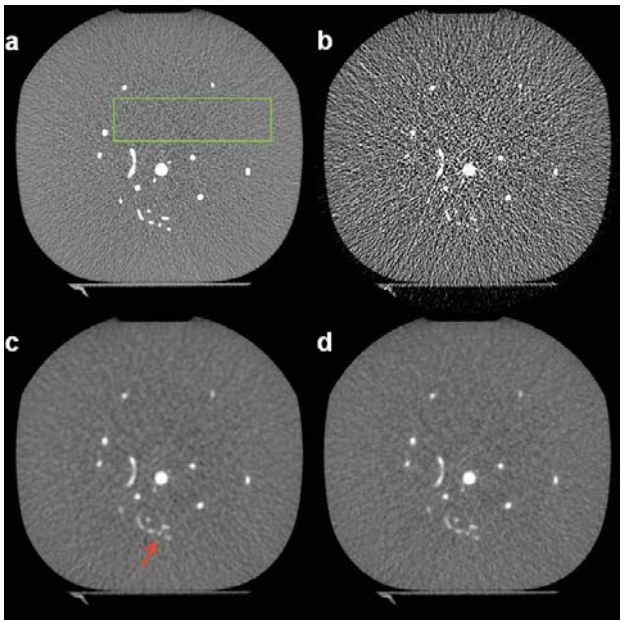


Fig. 1. Transaxial slice through the vessel phantom, level/window = 100/1500 HU: non-filtered normal dose (a), non-filtered 10% dose (b), 10% dose with Gaussian filtering (c), 10% dose with hTV filtering (d).

A. Vessel phantom

The vessel phantom consists of aluminium wires of different diameters in an almost cylindrical water background. The phantom has been scanned twice using 300 projections on 200 degree, once with standard clinical settings and once with only 10% of the X-ray dose by lowering the tube output. A transaxial slice through the non-filtered reconstruction acquired at normal dose is shown for comparison in Fig. 1a. A MIP view is shown in the left image of Fig. 2. The low dose acquisition has been filtered with the isotropic Gaussian and the TV based method. The filter threshold t has been set to 75% of the direct radiation for both filters. Reconstructions from the acquisition at 10% dose are shown without filtering,

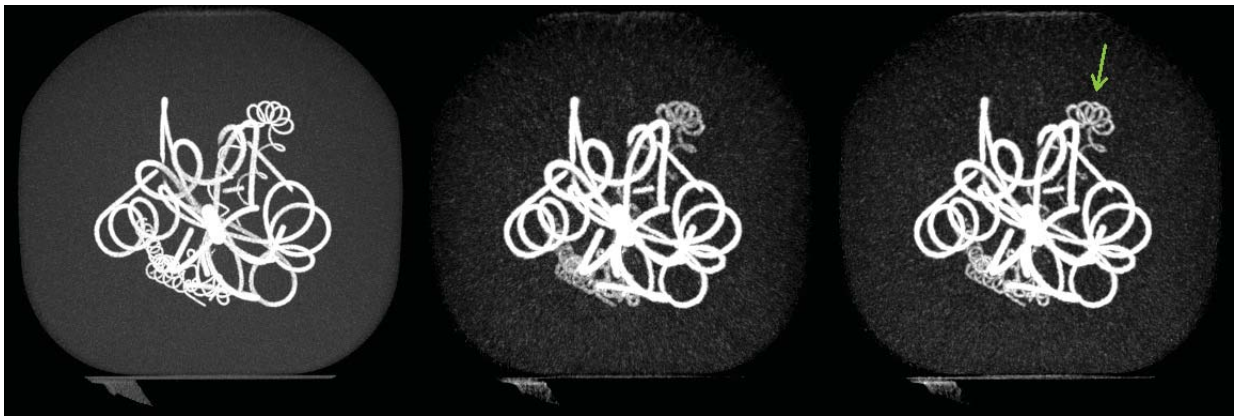


Fig. 2. Maximum intensity projection (MIP) in caudo-cranial direction of the vessel phantom: standard dose, L/W=900/2000 HU (left), Gaussian filtered low dose (middle), hTV filtered low dose (right), both L/W=300/500 HU.

Gaussian filtering and hTV filtering in Fig. 1b-d, respectively. The noise standard deviation has been measured for the low dose reconstructions in the green area indicated in Fig. 1a as 354 HU for the unprocessed data and 49 HU for both adaptive filters. A greater loss in spatial resolution can be observed for the Gaussian filtering compared to hTV filtering as indicated by the red arrow. This effect is better visualized by comparing the caudo-cranial MIPs of the filtered low dose reconstructions with the MIP of the non-filtered standard dose reconstruction (see Fig. 2). Especially the small vessels are better preserved by TV based filtering as indicated by the green arrow.

B. Clinical data

A rotational XperCT angiography of the liver with 308 projections on 200 degree has been processed with the adaptive filtering. The effect of Gaussian and hTV filtering on a single projection compared to non-filtered processing is shown in Fig. 3. The catheter and the proximal liver artery are slightly blurred by the isotropic Gaussian filter (red arrows) and better retained by the hTV filter (green arrows).

These projections from Fig. 3 have been used for the reconstruction of the corresponding volumes presented in Figs. 4,5. The noise standard deviation is measured in the homogeneous area highlighted by the rectangle (see blue arrow in Fig. 4). The noise is 44 HU for the non-filtered processing and 29 HU for both filtering methods. The reduced noise in the hTV filtered MIP allows better visualization of small vessels (see green arrows in Fig. 4), while the Gaussian filter deteriorates especially the visibility of vessels oriented in lateral direction that corresponds to the direction of strongest attenuation (see

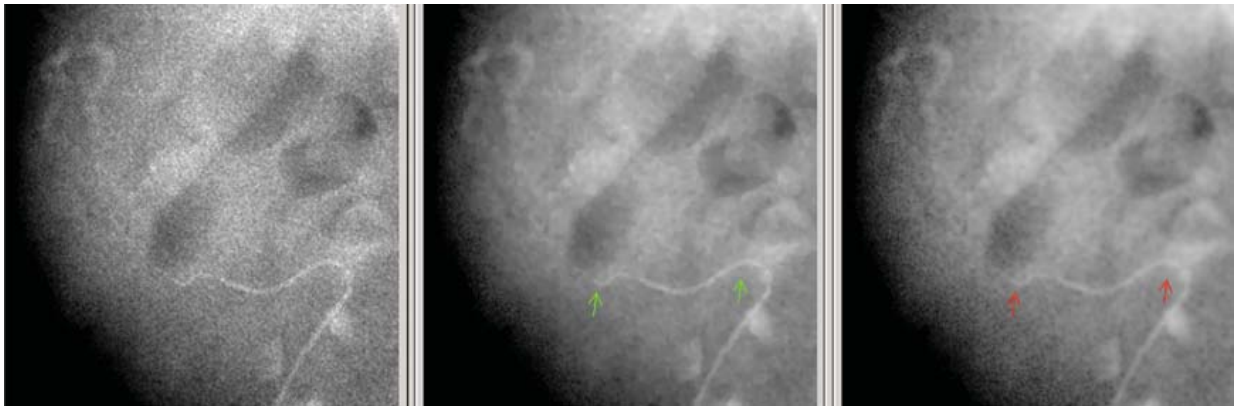


Fig. 3. Detail of a lateral projection of the liver angiography shown in Fig. 4: non-filtered (left), hTV filtering (middle), adaptive Gaussian filtering (right).

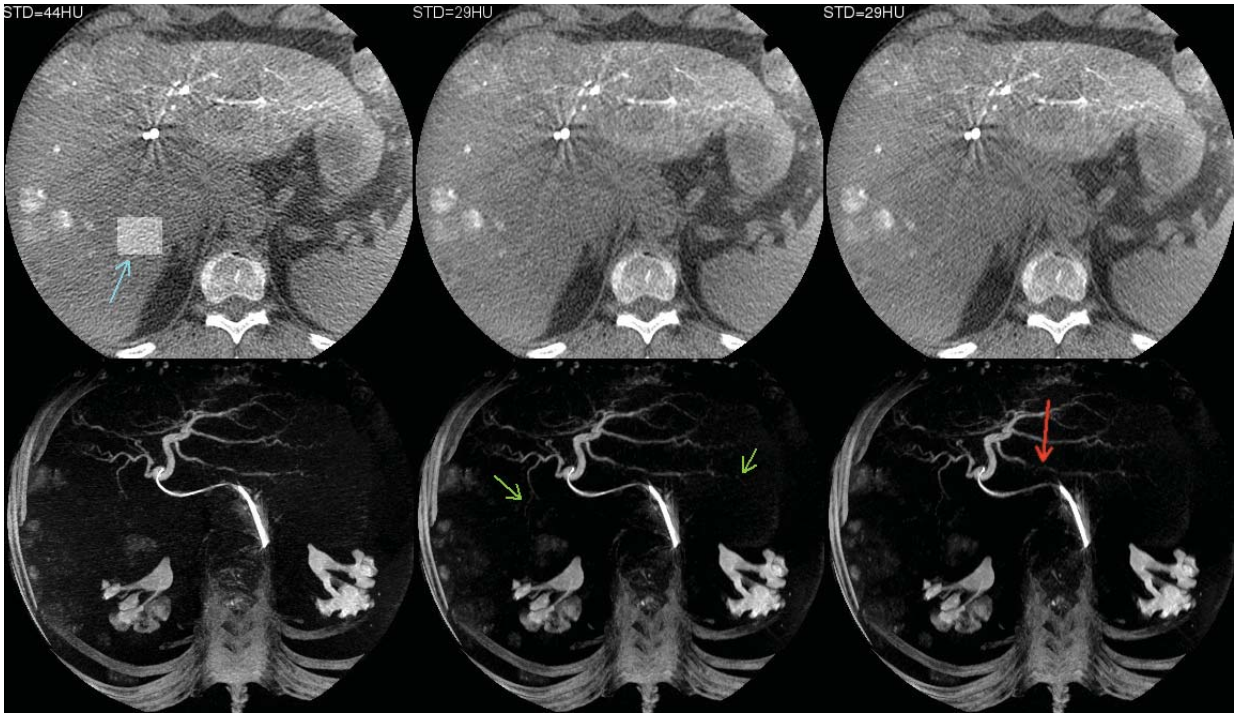


Fig. 4. Example of an angiographic liver scan, non-filtered (left column), hTV filtering (middle column), adaptive Gaussian filtering (right column), transaxial slice with 5 mm thickness, level/window 100/500 HU (top row), caudo-cranial MIP with level/window 1000/1500 HU (bottom row).

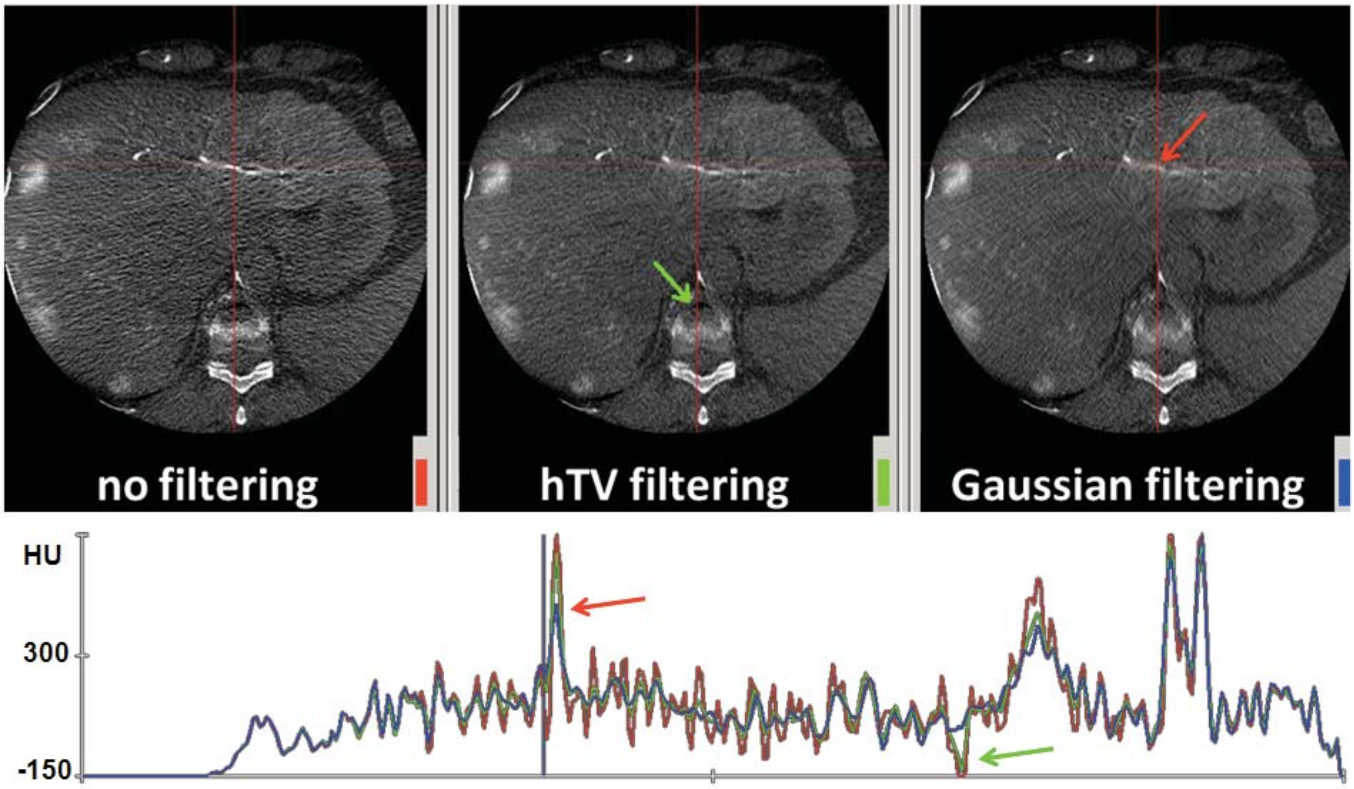


Fig. 5. Example of an angiographic liver scan with profiles, non-filtered (left, red profile), hTV filtering (middle, green profile), adaptive Gaussian filtering (right, blue profile), transaxial slice with 0.65 mm thickness, level/window 300/900 HU.

red arrow in Fig. 4). This vessel is also shown in the transaxial slice in Fig. 5 and marked there by a red arrow as well. The profiles show the noise removal capabilities and a better preservation of spatial resolution for hTV filtering compared to Gaussian filtering.

IV. Summary and discussion

We presented a TV based projection filter for adaptive filtering of rotational angiographic projections and compared the performance to well-known isotropic adaptive Gaussian filtering. For the phantom data the visibility of small vessels is better preserved with the hTV filter compared to Gaussian filtering at equal background noise level. This behavior is confirmed on a clinical liver angiography data set.

Acknowledgment

The authors thank K. Brown and S. Zabic, Philips Healthcare, USA for support and discussions during algorithm implementation as well as M. Lin and J. Geschwind, Johns Hopkins University School of Medicine, USA, for providing the clinical data.

References

[1] KM. Brown, S. Zabic, and T. Koehler, "Comparison of ML Iterative Reconstruction and TV-Minimization for Noise Reduction in CT Images," 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 443-446, 2011.

[2] T. Chan, J. Shen, *Image Processing and Analysis*, Chapter 4.5, "Variational Denoising based on the bounded variation image model," 174-191, 2005.

[3] L. Feldkamp, L. Davis, J. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A, Vol 1, 612-619. 1984.

[4] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," Med. Phys. Vol. 25(11), 2139-47, 1998.

[5] M. Kachelrieß, O. Watzke, and WA. Kalender, "Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT," Med. Phys. Vol. 28(4), 475-490, 2001.

[6] A. Manduca, L. Yu, JD. Trzasko, N. Khaylova, JM. Kofler, CM. McCollough, and JG. Fletcher, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," Med. Phys. Vol. 36(11), 4911-19, 2009.

[7] A. Maier, L. Wigström, HG. Hofmann, J. Hornegger, L. Zhu, N. Strobel, R. Fahrig, "Three-dimensional anisotropic adaptive filtering of projection data for noise reduction in cone beam CT," Med. Phys. Vol. 38(11), 5896-5909, 2011.

[8] L. Rudin, S. Osher, E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D, Vol.60, 259-268, 1992.

[9] E. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol., Vol. 53, 4777–4807, 2008.

[10] C.R. Vogel, "Computational Methods for Inverse Problems," SIAM, 2002.

[11] Z. Yang, AA. Zamyatin, and N. Akino, "Effective Data-domain Noise and Streak Reduction for X-Ray CT," 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 290-293, 2011.

[12] R Zanella, P Boccacci, L Zanni, and M Bertero, "Efficient gradient projection methods for edge-preserving removal of Poisson noise," Inverse Problems Vol. 25, 1-24, 2009

CT Reconstruction from Few-Views by Higher Order Adaptive Weighted Total Variation Minimization

Maurice Debatin, Dzmitry Stsepankou, Jürgen Hesser

***Abstract*— Dose reduction in X-ray Computed Tomography (CT) is of high practical relevance. Compressed Sensing allows for efficient under-sampling while still achieving an acceptable image quality. Especially Total Variation (TV) regularization obtains accurate, robust and stable results. However, it often suffers from the loss of fine structures and stair-casing artifacts.**

In order to overcome these limitations, we propose a generalization of TV by higher order derivatives. We demonstrate in this paper that both stair-casing and the loss of small structures in TV-based iterative tomographic reconstructions can be overcome.

***Index Terms*— CBCT, higher order derivatives, Anisotropic weighted Total Variation, low-dose**

I. Introduction

Cells are highly sensitive to ionizing radiation. Especially for CT, the observed dose is in the range of 20 mSv, i.e. one order of magnitude higher than the annual dose. Hence, dose reduction in diagnostic and therapeutic imaging is of substantial interest. In image-guided radiation therapy, where cone-beam CT is used, daily CT imaging could accumulate to a dose that is comparable to the biologically relevant dose for tumor control. This underlines the substantial interest in developing low-dose CT regimes.

A promising strategy to reduce the dose is to under-sample the target. However, using only a portion of the available dataset introduces artifacts in analytic reconstruction algorithms like filtered backprojection [1], although it is still the standard in modern CTs [2].

Iterative algorithms like Maximum Likelihood Expectation Maximization (ML-EM) [3,4,37] present an alternative to solve the data model when only a limited amount of measurements are available. Compressed Sensing (CS) [5] like Total Variation (TV) [8] partially overcomes the problem of data insufficiency where a series of results demonstrate an accurate and stable signal recovery from a highly limited number of observations [6,7,9-11]. The performance of TV was evaluated [12-15] and its clinical utility approved [16]. It can suppress noise and under-sampling artifacts and at the same time it preserves prominent edges. Despite the success of TV [17-22], there are considerable drawbacks. Among them are slow convergence rates that are accelerated by GPU implementations [23] or optimized versions of TV [24]. Anisotropic formulations of Total Variation (either by gradient redefinition [25], parameter-weighted [26] or Gaussian-like function-weighted formulations [27, 28] of TV) can overcome the loss of fine structures. However, the stair-casing effect withstands all these further developments in TV and in recent reformulations as Adaptive weighted Total Variation (AwTV) [28]. A promising strategy uses higher order derivatives [29-31,38] for the discretization of the gradient.

In our work, we will show that a combination of first order AwTV and second order TV overcomes the stair-casing artifact while keeping the advantageous features of AwTV, namely high spatial resolution and robustness to noise.

Manuscript received February 2, 2013, in final form April 4, 2013. This work was supported in part by BMWF and DFG.

M. Debatin, D. Stsepankou, J. Hesser are with the Department of Radiation Oncology, Mannheim Medical Centre, University of Heidelberg, Mannheim e-mail:{Maurice.Debatin,Dzmitry.Stspenakou,Juergen.Hesser}@medma.uni-heidelberg.de.

II. Material and Methods

A. The CT imaging model

The general acquisition paradigm in X-Ray tomographic imaging reads:

$$g = M \cdot f \qquad (1)$$

where g are the projection data, M is the system matrix and f are the absorption coefficients, f_i being the absorption coefficient for voxel i. The approximated measured projection data is considered as a random variable g with a Poisson density distribution and the measured value $\tilde{g}$ is a realization of g:

$$\tilde{g} \approx POISSON(\textstyle\sum_i m_i \cdot f_i) \qquad (2)$$

This approximation is valid after proper Noise Equivalent Counts (NEC) scaling [32]. For implementation details, see Stsepankou et al. [15]. The statistical bias-free maximum likelihood estimate of f_0 minimizes Csiszar's I-divergence [33]. The optimization function for the maximum *a posteriori* approach thus reads:

$$f_0 = argmin_f \left(I\left(\frac{M \cdot f}{\tilde{g}}\right) + \beta \cdot \psi(f) \right), \qquad (3)$$

$$\text{where } I\left(\frac{a}{b}\right) = \textstyle\sum_i b_i \cdot ln\left(\frac{b_i}{a_i}\right) - b_i + a_i$$

Prior knowledge can be incorporated into (3) by using a sparsity enforcing regularization function $\psi(f)$.

The image restoration model based on TV [8] is a strong tool to smooth noise and at the same time to keep prominent edges. Moreover, it tends to preserve the monotonicity of neighboring pixel values in the absence of noise. The TV gradient magnitude can be approximated by left sided differences and defined in 2 dimensions as:

$$\|\nabla f\|_1 = \qquad (4)$$

$$\sum_{x,y} |\nabla f_{x,y}| \approx \sum_{x,y} \sqrt{\left(\mathrm{f}_{x,y} - f_{x-1,y}\right)^2 + \left(\mathrm{f}_{x,y} - f_{x,y-1}\right)^2 + \varepsilon^2}$$

$\mathcal{E}$ is a small parameter usually set to $\mathcal{E} = 0.01$ in order to ensure differentiability at zero. x and y are the image dimension indices. A definition of a three-dimensional TV can be derived analogously.

In the presence of noise, TV cannot always guarantee preserving the monotonicity of neighboring pixel values. As a consequence, stair-casing artifacts occur, which means that supposedly "true" edges are in fact nothing else but large gradients which are misidentified because of high noise contribution. Furthermore, TV cannot smooth noise perpendicular to the edge, causing the loss of fine structures and low resolution.

Up to a certain noise level, the image restoration model based on AwTV [28] can differentiate between discontinuous components related to prominent edges and regions that are well approximated by moderate but nearly constant slopes. The AwTV gradient magnitude can be approximated in 2 dimensions by left sided differences and reads:

$$\|\nabla^w f\|_1 = \sum_{x,y} |\nabla^w f_{x,y}| \approx \qquad (5)$$
$$\sum_{x,y} \sqrt{\left(\mathrm{f}_{x,y} - f_{x-1,y}\right)^2 \cdot w_{f_{x,y},f_{x-1,y}} + \left(\mathrm{f}_{x,y} - f_{x,y-1}\right)^2 \cdot w_{f_{x,y},f_{x,y-1}} + \mathcal{E}^2},$$
$$\text{where } w_{a,b} = e^{-\frac{1}{2}\left(\frac{a-b}{\delta}\right)^2} \text{ and } \mathcal{E} = 0.01$$

The weighting function $w_{a,b}$ separates noise and edges up to a certain noise level. The separation is controlled by parameter δ. When δ is small, larger gradients are not penalized because $w_{a,b} \approx 0$. Consequently, the edges can be restored well since the contribution of regularization is low at such regions. With a large δ, the AwTV is close to TV in its behavior.

In a low-noise scenario, AwTV tends to reduce the staircase effect. Furthermore, it can increase the spatial resolution in contrast to TV, since the smoothing can be stopped at large gradients. However, if the noise surpasses a certain threshold, the AwTV model would still prefer stair casing over ramps, especially in areas where the underlying image is homogeneous.

A potential way to tackle this problem is to use higher order derivatives in order to more penalize jumps at such regions.

Intuitively consider a discrete step function at a height of $h = 1$. The first order derivative of this function is $\frac{1}{h}$ and the second order derivative is $\frac{1}{h^2} \gg \frac{1}{h}$, when $h \approx 0$. Consequently, higher order derivatives prefer smooth (linear) transitions over jumps.

We therefore propose a new functional which can distinguish jumps (prominent edges) from smooth transitions (homogeneous areas) even in the presence of high noise. Such a functional can be designed by combining first and second order derivatives [30]. The current image estimate f is decomposed in the sum of a function that contains the jumps and a smooth function. We propose to use our new regularization function $\psi(f)$ in (3) in order to address the aforementioned problems caused by TV and AwTV, which can be defined in 2 dimensions as:

$$(1-\lambda) \cdot \|\nabla^w f\|_1 + \lambda \cdot \|\Delta f\|_1, \text{ where} \qquad (6)$$
$$\|\Delta f\|_1 \approx$$
$$\sum_{x,y} \sqrt{\left(f_{x,y} + 2 \cdot f_{x-1,y} - f_{x-2,y}\right)^2 + \left(f_{x,y} + 2 \cdot f_{x,y-1} - f_{x,y-2}\right)^2 + \mathcal{E}^2},$$

$\mathcal{E} = 0.01$ and $\|\nabla^w f\|$ is defined in (5). x and y are the image indices. The first part of (6) is measured using AwTV since this functional has the best edge-preserving properties. The second part is measured using the higher order TV, which is able to suppress the stair-cases, since large gradients are more penalized. The ration is controlled by parameter λ.

It is the task of the minimization problem to correctly handle the precise decomposition of both parts. We therefore let the model decide for itself where to apply AwTV and where to apply TV².

B. *The reconstruction algorithm*

Maximum Likelihood Expectation Maximization (ML-EM) [3] is used to solve the minimization functional (3). Regularization is included by the One Step Late-approach (OSL) [34-36]. We abbreviate the identification of the reconstruction algorithm using the different regularization functionals TV, AwTV or AwTV+TV² by ML-EM-{TV, AwTV, AwTV+TV²}.

Forward and backward projection are the most computationally challenging operations which we mapped on a Graphics Processing Unit (GPU). Backward projection is voxel driven with bilinear interpolation on the projections. Forward projection is sample-based and uses texture mapping with trilinear interpolation. We simulate real clinical CT datasets by reconstructing the available data using the standard FDK algorithm.

III. RESULTS

The Catphan 600 phantom was scanned on an Elekta Synergy KV CBCT (XVI) system with a tube current of 20 mAs and reconstructed using all available 400 projection measurements and the FDK algorithm [1]. Then, 80 forward projections were simulated using a scanner angle interval of 5 degrees. A phantom similar to the digital **N**on-uniform rational basis spline (**N**URBS) based **Ca**rdiac **T**orso (NCAT) phantom [39] was reconstructed in a noise-free and scatter-free scenario with simulated settings and a geometry setting of an Elekta Synergy KV CBCT (XVI). 40 projections out of 400 available measurements were taken for the reconstruction with a scanner angle increment of 10 degrees. The projections of the NCAT phantom were simulated with a detector size of 256 × 256 pixels and a pixel size of 1.0 mm × 1.0 mm. All reconstruction results are of the resolution 256 × 256 × 256 voxels with the voxel size chosen as 1.0 mm × 1.0 mm × 1.0 mm. ML-EM-TV (4) and ML-EM-AwTV (5) were selected for the regularization and compared to the results obtained by our new approach ML-EM-AwTV+TV² (6). We used the following parameter settings:

The regularization parameter β was set to $\beta = 0.03$ in fig. 1 and $\beta = 0.035$ in fig. 2. δ was set to $5 \cdot 10^{-6}$ for both regularization functions (5) and (6). In (6), λ is controlling the ratio between first and second order derivatives and must lie in the interval [0, 1]. For our experiments, $\lambda = 0.17$ in fig. 1 and $\lambda = 0.25$ in fig. 2 produced the best results based on the quality metric inspection in table 1 and table 2.

In order to assess the non-homogeneity caused by e.g. stair-cases in the reconstruction results we introduce the **N**on-**U**niformity **E**rror **I**ndex (NUEI) as a reliable measure. The NUEI is computed by cutting out a Region Of Interest (ROI) at a location where the true image is known to be constant. The ROI can be of arbitrary length and width but should cover all available slices in depth. Then, the 3-dimensional Pixel Variance Map (PVM) can identify jumps and inhomogenities by computing for each pixel its $2 \times 2 \times 2$ neighborhood

variance. Then, the Standard Deviation (SD) of the PVM can be used to numerically determine how homogeneous the results are. A value of 0 refers to a completely homogeneous result. To summarize this, the NUEI can be computed by:

$NUEI = \frac{\sum_i SD(PVM(ROI_i)}{i}$, where i is the number of ROIs.

The results in fig. 1-2 and table 1-2 show that using AwTV+TV² as regularization, the stair-casing artifact can be suppressed. Our proposed method is up to 1.57 times more homogeneous than AwTV and up to 7.7 times more homogeneous than TV. For the visual assessment of the stair-cases we compute the contour map (level lines). In fig. 1-2 it can be clearly seen that for our proposed regularization method (6) in homogeneous areas the level lines of the contour map are all spaced far apart in contrast to the results produced by (4) and (5). There, the level lines which are lying side by side or overlap indicate an edge, which is undesired.

Intuitively, the assessment of spatial resolution bases on the ability to differentiate line-pairs from the background in fig. 2. Therefore, after normalization to 1, the pixel-intensities of the line (peaks) and the background (valleys) can be measured and their ratio can be determined by computing the Euclidean distance. A value of 1 would correspond to the highest possible resolution. We have examined line-pair number 3. Table 2 shows that ML-EM-AwTV and ML-EM-AwTV+TV² have a resolution that is nearly the same and both are approximately 26 % higher compared to ML-EM-TV. One can thus conclude that there is practically no tradeoff regarding resolution when AwTV+TV² is used as regularization as if solely first order derivatives had been used, e.g. AwTV. However, for both regularization functions, TV and AwTV, strong stair-case artifacts make the results look unnaturally in contrast to our proposed method AwTV+TV².

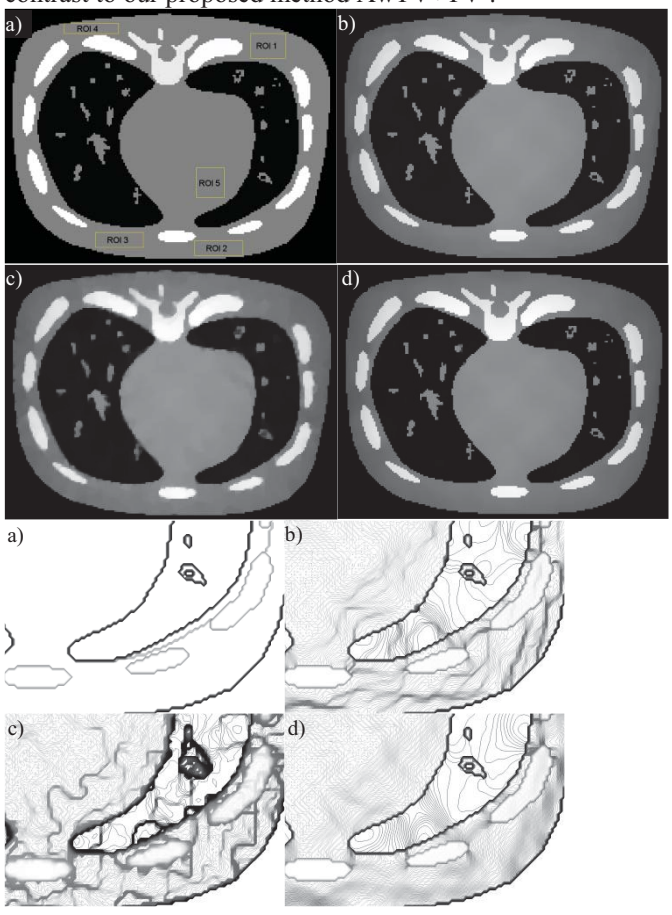


Figure 1 Results showing the reconstruction of the three dimensional NCAT phantom. 40 projections were used for the reconstruction. All figures in the first and second rows have gray values in [0,1], the corresponding zoomed versions show a contour map (level lines) of the reconstructions. a) the ground truth b) ML-EM-AwTV, c) ML-EM-TV, d) our new approach: ML-EM-AwTV+TV². Note that for our proposed method ML-EM-AwTV+TV² in homogeneous areas, the level lines of the contour map are all spaced far apart in contrast to the other results in b) and c). There, the level lines which are overlapping or lying directly side by side indicate an edge, which is undesired.

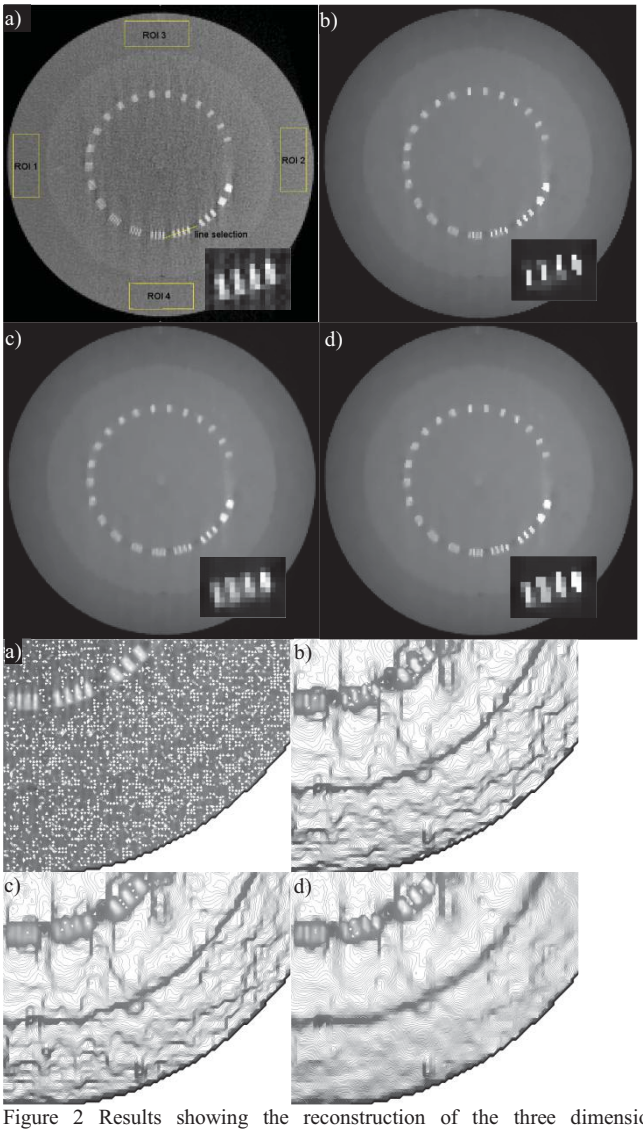


Figure 2 Results showing the reconstruction of the three dimensional simulated Catphan 600 phantom. The ML-EM algorithm and 80 projections were used for the reconstruction. All figures in the first and second rows have gray values in [0,1], the corresponding zoomed versions show a contour map (level lines) of the reconstructions. a) the ground truth b) ML-EM-AwTV, c) ML-EM-TV, d) our new approach: ML-EM-AwTV+TV². It can be clearly seen that our approach d) can drastically reduce the stair-casing effect. In homogeneous areas most of the level lines lie far apart, not so for the results in b) and c) where the level lines which are very close to one another or even overlap indicate an edge, which is unwanted.

TABLE I
NON-UNIFORMITY ERROR INDEX

	ML-EM-AwTV+TV²		ML-EM-AwTV		ML-EM-TV	
	CATPHAN	NCAT	CATPHAN	NCAT	CATPHAN	NCAT
NUEI	**1.58**	**0.94**	2.18	1.48	2.12	7.26

The Non-Uniformity Error Index of the Region Of Interests selected in fig.1 and fig.2. ML-EM-AwTV+TV² has the smallest NUEI. It is therefore the most homogeneous result. All values have been scaled by a factor of 10^6.

TABLE 2
SPATIAL RESOLUTION OF LINE PAIR NO. 3 CATPHAN RESOLUTION SLICE

Resolution ratio [0,1]	ML-EM-AwTV+TV²	ML-EM-AwTV	ML-EM-TV
mean	0.77	**0.78**	0.61

The mean ratio of function-value based peaks and valleys in the normalized line pair number 3 of the Catphan 600 phantom. Note that ML-EM-AwTV+TV² and ML-EM-AwTV have nearly the same spatial resolution.

IV. DISCUSSION AND CONCLUSION

We have shown that higher order derivatives used in the regularization function of CT reconstruction algorithms can suppress the stair-casing effect. Our proposed method is up to 1.57 times more homogeneous than AwTV and up to 7.7 times more homogeneous than TV. In addition to that, combined with first order AwTV, prominent edges, small details and structures can be restored. The spatial resolution for both regularization methods AwTV and our proposed one AwTV+TV² is approximately 26 % higher compared to TV. Consequently, one can conclude that there is no tradeoff regarding spatial resolution when first and second order derivatives are used in combination as regularization in x-ray Computed Tomography. The examination of the contour map indicates that our method produces the best result, since all level lines in constant regions lie far apart in contrast to the other methods; where level lines which are close to each other or even overlap indicate an edge, which is undesired.

REFERENCES

[1] Feldkamp, L. A., L. C. Davis, and J. W. Kress. "Practical cone-beam algorithm." JOSA A 1.6 (1984): 612-619.

[2] Pan, Xiaochuan, Emil Y. Sidky, and Michael Vannier. "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?." Inverse problems 25.12 (2009): 123009.

[3] Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society. Series B (Methodological) (1977): 1-38.

[4] Fessler, Jeffrey A., and Alfred O. Hero. "Space-alternating generalized expectation-maximization algorithm." Signal Processing, IEEE Transactions on 42.10 (1994): 2664-2677.

[5] Donoho, David Leigh. "Compressed sensing." Information Theory, IEEE Transactions on 52.4 (2006): 1289-1306.

[6] Candes, Emmanuel J., and Terence Tao. "Near-optimal signal recovery from random projections: Universal encoding strategies?." Information Theory, IEEE Transactions on 52.12 (2006): 5406-5425.

[7] Candes, Emmanuel J., Justin K. Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements." Communications on pure and applied mathematics 59.8 (2006): 1207-1223.

[8] Rudin, Leonid I., Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms." Physica D: Nonlinear Phenomena 60.1 (1992): 259-268.

[9] Sidky, Emil Y., Chien-Min Kao, and Xiaochuan Pan. "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT." Journal of X-ray Science and Technology 14.2 (2006): 119-139.

[10] Sidky, Emil Y., and Xiaochuan Pan. "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." Physics in medicine and biology 53.17 (2008): 4777.

[11] Sidky, Emil Y., and Xiaochuan Pan. "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." Physics in medicine and biology 53.17 (2008): 4777.

[12] Tang, Jie, Brian E. Nett, and Guang-Hong Chen. "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms." Physics in medicine and biology 54.19 (2009): 5781.

[13] Bian, Junguo, et al. "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT." Physics in Medicine and Biology 55.22 (2010): 6575.

[14] Kim, Dae-Hong, et al. "Evaluation of Reconstructed Images on the micro-CT system using Total Variation Minimization", (NSS/MIC), 2012 IEEE. IEEE, 2012.

[15] Stsepankou, D., et al. "Evaluation of robustness of maximum likelihood cone-beam CT reconstruction with total variation regularization." Physics in Medicine and Biology 57.19 (2012): 5955.

[16] Tian, Zhen, et al. "Low-dose CT reconstruction via edge-preserving total variation regularization." Physics in Medicine and Biology 56.18 (2011): 5949.

[17] Chen, Guang-Hong, Jie Tang, and Shuai Leng. "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets." Medical physics 35.2 (2008): 660.

[18] Choi, Kihwan, et al. "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method." Medical physics 37.9 (2010): 5113.

[19] LaRoque, Samuel J., Emil Y. Sidky, and Xiaochuan Pan. "Accurate image reconstruction from few-view and limited-angle data in diffraction tomography." JOSA A 25.7 (2008): 1772-1782.

[20] Kinouchi Shoko, et al. "Total variation minimization for in-beam PET image reconstruction", (NSS/MIC), 2012 IEEE. IEEE, 2012.

[21] Chou Hung-Yi, "Time-of-Flight Image Reconstruction with TV Minimization Constraint for a Dual-Head Small Animal PET System," (NSS/MIC), 2012 IEEE. IEEE, 2012.

[22] Nett, Brian, et al. "Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (PICCS) on a C-arm system." Proceedings-Society of Photo-Optical Instrumentation Engineers. Vol. 6913. NIH Public Access, 2008.

[23] Jia, Xun, et al. "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation." Medical Physics 37.4 (2010): 1757-1760.

[24] Sidky, Emil Y., Jakob H. Jørgensen, and Xiaochuan Pan. "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm." Physics in Medicine and Biology 57.10 (2012): 3065.

[25] Debatin, Maurice, et al. "CT Reconstruction from Few-Views by Anisotropic Total Variation Minimization", (NSS/MIC), 2012 IEEE. IEEE, 2012.

[26] Jin, Xin, et al. "Anisotropic total variation for limited-angle CT reconstruction." Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE. IEEE, 2010.

[27] Xu, Qiong, et al. "CT reconstruction based on improved total variation minimization." Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE. IEEE, 2010.

[28] Liu, Yan, et al. "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction." Physics in medicine and biology 57.23 (2012): 7923.

[29] Bredies, Kristian, Karl Kunisch, and Thomas Pock. "Total generalized variation." SIAM Journal on Imaging Sciences 3.3 (2010): 492-526.

[30] Chambolle, Antonin, and Pierre-Louis Lions. "Image recovery via total variation minimization and related problems." Numerische Mathematik 76.2 (1997): 167-188.

[31] Setzer, S., G. Steidl, and T. Teuber. "Infimal convolution regularizations with discrete l1-type functionals." Comm. Math. Sci 9 (2011): 797-872.

[32] Nuyts, Johan, Christian Michel, and Patrick Dupont. "Maximum-likelihood expectation-maximization reconstruction of sinograms with arbitrary noise distribution using NEC-transformations." Medical Imaging, IEEE Transactions on 20.5 (2001): 365-375.

[33] Snyder, Donald L., et al. "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography." Medical Imaging, IEEE Transactions on 6.3 (1987): 228-238.

[34] Green, Peter J. "Bayesian reconstructions from emission tomography data using a modified EM algorithm." Medical Imaging, IEEE Transactions on 9.1 (1990): 84-93.

[35] Green, Peter J. "On use of the EM for penalized likelihood estimation." Journal of the Royal Statistical Society. Series B (Methodological) (1990): 443-452.

[36] Panin, V. Y., G. L. Zeng, and G. T. Gullberg. "Total variation regulated EM algorithm [SPECT reconstruction]." Nuclear Science, IEEE Transactions on 46.6 (1999): 2202-2210.

[37] Hudson, H. Malcolm, and Richard S. Larkin. "Accelerated image reconstruction using ordered subsets of projection data." Medical Imaging, IEEE Transactions on 13.4 (1994): 601-609.

[38] Dobson, David C., and Fadil Santosa. "Recovery of blocky images from noisy and blurred data." SIAM Journal on Applied Mathematics 56.4 (1996): 1181-1198.

[39] Segars, William Paul, et al. "Development and application of the new dynamic Nurbs-based Cardiac-Torso (NCAT) phantom." Doctoral Dissertation, The University of North Carolina, Chapel Hill (2001).

Cylindrical coordinate representation for statistical 3D CT reconstruction

Yves Goussard*†, Mahsa Golkar*, Adrien Wagner* and Matthieu Voorons*
*Institut de génie biomédical, École Polytechnique de Montréal, Montréal, Québec, Canada
†Centre de recherches de l'Hôpital du Sacré-Cœur, 5400, Boul. Gouin Ouest, Montréal, Québec, Canada

***Abstract*—Iterative statistical reconstruction methods appear as an increasingly interesting alternative to conventional, analytical techniques. However, performance improvement in terms of amount of computation and memory footprint is a major requirement for wider adoption of the approach. This communication deals with one element that has a strong impact on the overall performance of 3D iterative reconstruction techniques, *i.e.,* representation of the object, and its impacts on the structure of the projection matrix and on the implementation of the projection and backprojection operations. More specifically, we focus on representation of the object in a cylindrical coordinate system, which presents the advantage of considerably reducing the memory footprint of the projection matrix by taking advantage of a large number of rotational invariances. We investigate the properties of the representation with respect to implementation, computational efficiency, development of penalized likelihood reconstruction techniques and elements that can impact the convergence of such estimators.**

I. Introduction and problem statement

Iterative *algebraic* or *statistical* reconstruction methods (SRMs) are now recognized as useful alternatives to analytical methods for obtaining highly accurate computed tomography (CT) reconstruction results. However, widespread use of SRMs is hindered by their high computational load memory footprint, particularly in three-dimensional (3D) X-ray CT, due to the large size of the object and of the projection data. Whichever specific SRM is used, most of the computational effort is spent during projection and backprojection operations. Therefore, parsimonious representation of the projection matrix and efficient implementation of the left and right matrix-vector products involving it have a strong impact on the performance and usability of SRMs.

In conventional approaches, the whole object to be reconstructed is discretized on a regularly spaced Cartesian grid. Even if symmetry properties and rotational invariance are taken advantage of, simple and efficient manipulation and storage of the *whole* projection are almost impossible to achieve. The solution generally consists of computing a significant part of the entries of the projection matrix "on the fly", which clearly results in a loss of computational efficiency. In order to circumvent these difficulties, alternative approaches have been proposed. In *targeted CT*, it is assumed that high quality reconstruction is required only in a small region of interest (ROI) of the object. One may split the object into two separate regions: a coarse background and the high resolution ROI, each of them discretized on Cartesian grids with different stepsizes [1]; interesting results have been obtained, but computations remain heavy, and implementation and practical use present a significant level of difficulty. One may also discretize the object on an adaptive irregular mesh, with a high vertex density in the ROI and low vertex density in the background [2]. This raises questions about mesh construction, practical computation of the projection matrix and conditioning of the reconstruction problem.

In *rotation-based* approaches, the general idea is to use a *partial* projection operator limited to a single or a small number of projection angles and to rotate the object appropriately so as to perform the complete projection and backprojection operations. This clearly results in a dramatic reduction in the size of the projection matrix, particularly when the number of projections is large, as in X-ray CT. The object may be discretized on a regular Cartesian grid [3]; rotation of the object then requires interpolation, which is time and memory consuming and generates approximations. These approximations may make the projection and backprojection operators not exact transpose of each other, which generally hinders the convergence of SRMs. This explains why the approach has received a relatively limited attention. The object may also be discretized in a *sector-invariant* manner, each sector corresponding to the angular scope of the partial projection operator. Sector-wise rotation of the object corresponds to a simple circular shift of the vector holding the discretized object samples, and the complete projection set can thus be computed efficiently by repeated application of the partial projection operator to the circular-shifted object. In addition, the partial projection operator can be easily precomputed and stored. This approach has been used mainly in PET reconstruction, for which several sector tessellation schemes have been proposed [4], [5]. Unfortunately, the complexity of the tessellation schemes makes computation of the partial projection operator far from trivial, and does limit the flexibility of the approach.

In 3D X-ray CT, a similar idea has been proposed, in which the object is discretized on a regular cylindrical grid, the angular stepsize being equal to the difference between consecutive projection angles [6]. The limited projection operator is very small in size and easy to compute, due to the geometric simplicity of the discretization scheme. The approach retains the same computational advantages as the other sector-invariant approaches. However, using such a representation raises several questions that could affect the effectiveness of SRMs, among which (*i*) computational efficiency of projection and backprojection, particularly regarding parallelization; (*ii*) conditioning of the reconstruction problem, which could be poor due the variable size of the object samples; (*iii*) derivation of penalty functions consistent with the cylindrical discretization scheme; (*iv*) actual quality of reconstructions results. This communication provides answer elements to these questions.

II. Assumptions and reconstruction framework

We consider a 3D axial CT reconstruction problem in which the data formation model can be written as:

$$\boldsymbol{y} = \mathbf{P}\boldsymbol{\mu} + \boldsymbol{b} \tag{1}$$

Here, $\boldsymbol{y} \in \mathbb{R}^{MN}$, $\boldsymbol{\mu} \in \mathbb{R}^{KN}$ and $\boldsymbol{b} \in \mathbb{R}^{MN}$ respectively denote the vectors containing the complete set of projection data, the object samples and measurement noise. $\mathbf{P}$ represent the complete projection operator. It is assumed that the object is discretized onto a regular cylindrical grid with number of angular samples N equal to the number of regularly spaced projections. Scalars K and M represent the number of radial samples of the object and the number of data samples per projection angle, respectively. It is assumed that the samples in $\boldsymbol{y}$ and $\boldsymbol{\mu}$ are regularly arranged by increasing angle values.

Model (1) is a useful approximation to a wide class of transmission and emission tomography problems.

Assuming $\boldsymbol{b} \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Sigma}^{-1})$ and an exponential prior distribution of $\boldsymbol{\mu}$, MAP reconstruction corresponds to solving:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \; J(\boldsymbol{\mu}) \tag{2}$$

$$J(\boldsymbol{\mu}) = \tfrac{1}{2}\|\boldsymbol{y} - \mathbf{P}\boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}} + \lambda R(\boldsymbol{\mu}) \tag{3}$$

Matrix $\boldsymbol{\Sigma}$ is a diagonal matrix that can be used to account for the Poisson distribution of X-ray photon counts [7]. $\lambda R(\boldsymbol{\mu})$ is a regularization function that applies a L_2 (quadratic) or L_2L_1 (edge-preserving) penalty on $\boldsymbol{\mu}$ itself and/or on its first differences on first- or second-order neighborhoods. In the sequel, it is assumed that optimization problem is solved using an iterative, descent algorithm which requires at least one projection and one backprojection operation per iteration. A large fraction of commonly used SRMs do present such a characteristic, and evaluation of projections and backprojections then dominate the computations.

III. Projection and Backprojection

Recall that, under our assumptions, the number N of angular samples of $\boldsymbol{\mu}$ is equal to the number of equally spaced projections. Let $\boldsymbol{y}_n$ denote the projection data for projection angles $\theta_n \; ; \; 0 \leq n \leq N-1$ and let $\mathbf{P}_0$ represent the partial projection matrix for a reference projection angle that can be set to θ_0 without loss of generality. $\mathbf{P}_0$ can be evaluated easily using a ray-driven approach (see [6] for details), and one can easily account for ray thickness several thin rays per detector. The relatively small size and sparse structure of $\mathbf{P}_0$ make it easy to pre-compute and store.

Computation of $\boldsymbol{y}_n$ requires rotating $\boldsymbol{\mu}$ by angle $-\theta_n$ prior to applying operator $\mathbf{P}_0$; in the cylindrical coordinate used for discretizing $\boldsymbol{\mu}$, this rotation corresponds to the repeated application of $\mathbf{S}_{\boldsymbol{\mu}}^{-1}$ n times, where $\mathbf{S}_{\boldsymbol{\mu}}$ is a simple, K-sample circular shift operator corresponding to the rotation of angle θ_1. Therefore, the complete projection operator $\mathbf{P}$ can be expressed as:

$$\mathbf{P} = \left[\mathbf{P}_0^{\mathrm{t}} \mid (\mathbf{P}_0\mathbf{S}_{\boldsymbol{\mu}}^{-1})^{\mathrm{t}} \mid \cdots \mid (\mathbf{P}_0\mathbf{S}_{\boldsymbol{\mu}}^{-N+1})^{\mathrm{t}}\right]^{\mathrm{t}} \tag{4}$$

Now, observe that computing $\mathbf{P}_0(\mathbf{S}_{\boldsymbol{\mu}}^{-n}\boldsymbol{\mu})$, *i.e.,* applying $\mathbf{P}_0$ to the Kn-sample, up circular-shifted version of $\boldsymbol{\mu}$ is equivalent to computing $(\mathbf{P}_0\mathbf{S}_{\boldsymbol{\mu}}^{-n})\boldsymbol{\mu}$, *i.e.,* applying the Kn-sample, right circular-shifted version of $\mathbf{P}_0$ to $\boldsymbol{\mu}$. In other words, $\mathbf{P}$ exhibits a (N, N) block-circulant structure with first block-row $\mathbf{P}_0$ given by:

$$\mathbf{P}_0 = [\mathbf{P}_{0,0} \mid \mathbf{P}_{0,1} \mid \cdots \mid \mathbf{P}_{0,N-1}] \tag{5}$$

where each $\mathbf{P}_{0,n} \; ; \; 0 \leq n \leq N-1$ is a (M, K) block.

The block-circulant structure of $\mathbf{P}$ lends itself to efficient implementation of projection and backprojection operations. For computing projections, the block-row decomposition of (4) is appropriate, since application of $\mathbf{S}_{\boldsymbol{\mu}}^{-n}$ involves no arithmetic computation and can be handled through index manipulations; furthermore, efficient representation of and products by $\mathbf{P}_0$ can be achieved by using one of several publicly available libraries. In addition, projections at different angles $\{\boldsymbol{y}_n\}$ can be computed independently from one another. This computational structure lends itself to easy parallelization, as access serialization must be imposed on constant vector $\boldsymbol{\mu}$ only. This makes general purpose parallelization libraries such as OpenMP® trivial to use.

For computing the backprojection of $\boldsymbol{y}$ onto a vector $\boldsymbol{\mu}^{(b)}$, it can easily be seen from (4) that:

$$\boldsymbol{\mu}^{(b)} = \sum_{n=0}^{N-1} \mathbf{S}_{\boldsymbol{\mu}}^{n}\mathbf{P}_0^{\mathrm{t}}\boldsymbol{y}_n \tag{6}$$

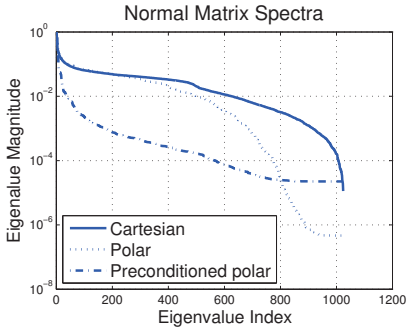


Fig. 1. Eigenvalue spectra of the normal matrix $\mathbf{P}^{\mathrm{t}}\mathbf{P} + \lambda\mathbf{I}$ in the Cartesian and cylindrical representations. 2D framework, $\lambda = 10^{-3}$.

where we have made use of the orthogonal structure of $\mathbf{S}_{\boldsymbol{\mu}}$. The computations in (6) are easy to implement, but their structure does not lend itself to parallel implementation because the accumulation into $\boldsymbol{\mu}^{(b)}$ must be serialized. However, the difficulty can be circumvented by using the block-circulant structure of $\mathbf{P}$; partitioning $\mathbf{P}$ in a block-column manner yields:

$$\mathbf{P} = \left[\mathbf{Q}_0 \mid \mathbf{S}_{\boldsymbol{y}}^{-1}\mathbf{Q}_0 \mid \cdots \mid \mathbf{S}_{\boldsymbol{y}}^{-N+1}\mathbf{Q}_0\right] \tag{7}$$

with

$$\mathbf{Q}_0 \; = \left[\mathbf{P}_{0,0}^{\mathrm{t}} \mid \mathbf{P}_{0,N-1}^{\mathrm{t}} \mid \mathbf{P}_{0,N-2}^{\mathrm{t}} \mid \cdots \mid \mathbf{P}_{0,1}^{\mathrm{t}}\right]^{\mathrm{t}} \tag{8}$$

where $\mathbf{S}_{\boldsymbol{y}}$ denotes the M-sample circular shift operator of appropriate size. (7) and (8) show that $\mathbf{P}^{\mathrm{t}}$ can be decomposed in the exact same manner as $\mathbf{P}$ in (4). Therefore, efficient computation and parallelization of backprojection operations can be achieved in the same way as for projections. Therefore, the above elements clearly indicate that the cylindrical representation does present interesting characteristics with respect to memory footprint and efficiency of projection and backprojection operations.

IV. Conditioning

Discretization of the object onto a regular cylindrical grid yields a variable voxel size. One may conjecture this will result in a deterioration of the conditioning of the reconstruction problem, and therefore in a slower convergence of SRMs. This point is confirmed experimentally by comparing the eigenvalue spectra of the normal matrices of the reconstruction problem when the projection matrix is expressed in Cartesian and cylindrical coordinates. A typical example of such spectra is presented in Fig. 1. It can be observed that, when the cylindrical representation is used, the eigenvalues are spread over a range significantly wider than with the Cartesian representation. In order to alleviate the difficulty, one may attempt to use preconditioning techniques. Compensation of the variable voxel size resulting from the cylindrical representation suggests the derivation of a diagonal preconditioner, either based on some heuristics related to the voxel size or derived from the computation of the diagonal elements of the normal matrix[1]. Fig. 1 also presents the spectrum of the cylindrical normal matrix preconditioned by the inverse of its diagonal elements. The spread of the eigenvalues is greatly reduced, which strongly suggests that diagonal preconditioning can alleviate the deterioration of the conditioning of the reconstruction problem associated with the cylindrical representation.

[1]For the cylindrical representation, these diagonal elements can be computed rather easily using (4).

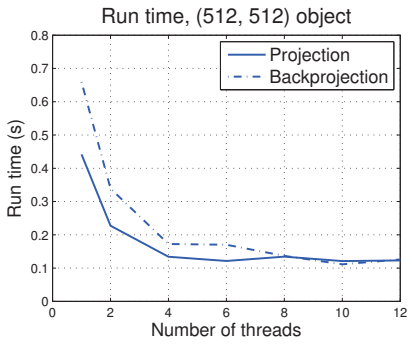


Fig. 2. Computation time of the projection and backprojection operations, as a function of the number of threads.

V. Penalty function in cylindrical coordinates

As indicated in Section II, regularization term $R(\boldsymbol{\mu})$ generally applies a penalty on first differences of $\boldsymbol{\mu}$. The rationale for this approach is to penalize some numerical approximation to the gradient of $\boldsymbol{\mu}$. When $\boldsymbol{\mu}$ is discretized onto a regular Cartesian grid, little difficulties occur since the distance between neighboring voxels is identical along each direction in which the differences are taken, and does not vary widely across directions. However, when $\boldsymbol{\mu}$ is discretized onto a cylindrical grid, the distance between neighboring voxels in some directions (*i.e.,* tangential) varies significantly, thereby giving a very strong weight to voxels located near the rotation axis and potentially affecting the conditioning of the reconstruction problem. For these reasons, it is unclear whether the first differences should be weighted by the inverse of the distance between voxels, which would be consistent with the physical meaning of the penalty term, or whether the first differences should be weighted in a uniform manner. This point will be investigated experimentally.

VI. Experimental results

A. Parallel implementation of projection and backprojection

The goal of these experiments was to assess the effectiveness of the simple parallelization scheme for the projection and backprojection operations outlined in Section III. The experiments were conducted in a 2D framework on realistic-size simulated data: $(512, 512)$ images, 512 equally spaced detectors, 1024 projections per rotation. The size of the polar grid used to discretize $\boldsymbol{\mu}$ was $(256, 1024)$ and covered a disk whose diameter was equal to the sidelength of the square Cartesian image. Sparse projection operators $\mathbf{P}_0$ and $\mathbf{Q}_0$ were computed and stored using an incremental CRS scheme [8]. Encoding of the sparse matrices and sparse matrix-dense vector products were programmed in C^{++} using Albert-Jan N. Yzelman's publicly available SparseLibrary[2]. Parallelization was performed in an elementary manner using the OpenMP® library.

Fig. 2 depicts the computation time of one projection and one backprojection averaged over 30 runs, as a function of the number of threads used, on an Intel multi-threaded 6-core processor running Linux. One can observe the typical behavior of simple parallelization schemes, which loose efficiency as the number of threads increases. However, these results were obtained at a very little cost and the reduction in computation time is significant with respect to a sequential implementation, which is encouraging for the potential effectiveness of more sophisticated parallelization schemes.

[2]Library and documentation available at http://people.cs.kuleuven.be/~albert-jan.yzelman/software.php.

B. Penalty function

As discussed in Section V, it is unclear whether the first differences should be weighted uniformly or in proportion to the inverse of the distance between neighboring voxels in the cylindrical coordinate system. The question was addressed in an experimental manner, through simulations performed in a 2D framework. Typical results are presented in Fig. 3: a Shepp-Logan numeric phantom was used with discretization and scanner geometric parameters identical to those used in Section VI-A, and Poisson noise was added to the projection data so as to obtain a global 30 dB signal-to-noise (SNR) ratio. Regularization function $R(\boldsymbol{\mu})$ consisted of a L_2 penalty applied to all first differences of $\boldsymbol{\mu}$ over second-order neighborhoods, with uniform and non-uniform weighting. Minimization of the resulting criteria was perform with a L-BFGS algorithm with a large number of iterations (*i.e.,* 500) in order to make sure that convergence was achieved. It can be observed that the uniformly weighted penalty function yields more accurate final reconstruction and faster convergence toward the actual attenuation map than the non-uniformly weighted penalty function. These kinds of results, which were observed in a variety of experimental conditions, strongly suggest that uniformly weighted penalty function should be preferred even when $\boldsymbol{\mu}$ is discretized in a cylindrical coordinate system.

C. Preconditioning and convergence

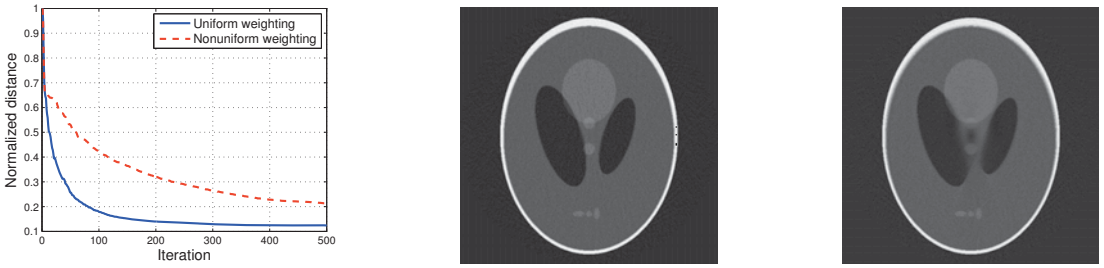
We now investigate the convergence of the proposed approach with respect to standard discretization on a Cartesian grid, and we also assess the efficiency of the diagonal preconditioning technique proposed in Section IV. The experimental conditions were identical to those used in Section VI-B. Uniform weighting was applied to the first differences in the L_2 penalty term; reconstructions were stopped after 100 iterations in order make the differences between methods appear more clearly. The results are presented in Fig. 4. As expected, the cylindrical approach without preconditioning presents a slower convergence than the standard Cartesian approach. However, a simple diagonal preconditioning based on the pixel size corrects for this deficiency, as no significant difference between the Cartesian and preconditioned cylindrical approaches can be observed. As similar results have been obtained with other types of penalty functions (*e.g.,*, edge-preserving L_2L_1), one can conclude that diagonal preconditioning compensates for the effect of variable-size voxels in the cylindrical representation.

D. Real data reconstruction

We now illustrate the use of the proposed approach for reconstruction of real 3D data. A section of the Catphan 600 phantom was scanned using a Siemens SOMATOM Sensation 16 tomograph in axial mode (SpineSeq L3-L4 protocol). The data were composed of 1160 projections of (672×12) samples each. The reconstructed volume was a (250×9) mm. cylinder sampled on a $256 \times 1160 \times 6$ polar grid. Multiple rays (9 per detector) were used to compute the projection operator, edge preserving L_2L_1 penalization was used in the regularization term and diagonal preconditioning was employed during reconstruction. A central slice of the reconstructed object, converted into Cartesian coordinates, is presented in Fig. 5. It does not exhibit any significant difference with reconstructions performed with standard SRMs, albeit being obtained with significantly lower reconstruction time and memory footprint.

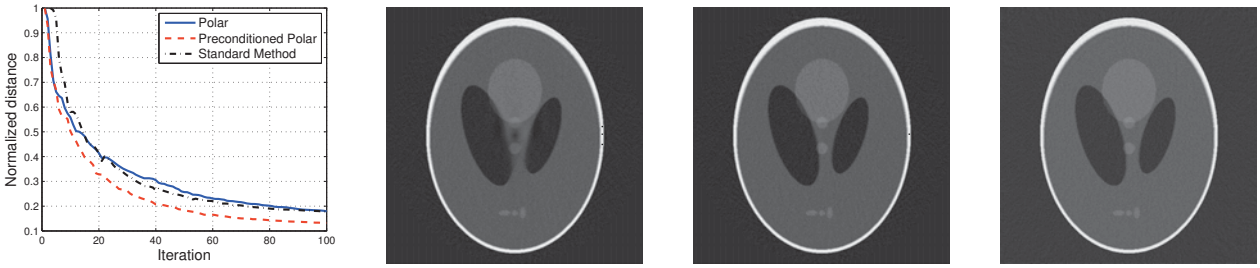
VII. Conclusion

Representing the object in a cylindrical coordinate system for 3D CT reconstruction presents appealing characteristics: low memory



(a) (b) (c)

Fig. 3. Comparison of uniform and non-uniform weighting of the first differences in the penalty function. (a) Relative L_2 distance to the actual object; (b) final reconstruction, uniform weighting; (c) final reconstruction, non-uniform weighting. Note the poor reconstruction in the center of (c).



(a) (b) (c) (d)

Fig. 4. Convergence of the proposed approach. (a) Relative L_2 distance to the actual object; (b) final reconstruction, cylindrical representation without preconditioning; (c) final reconstruction, cylindrical representation with diagonal preconditioning; (d) final reconstruction, standard approach.

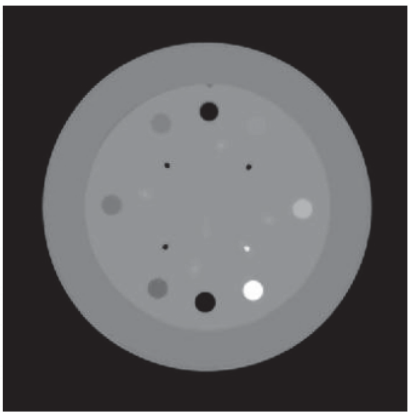


Fig. 5. Slice 3 of the $(512, 512)$ reconstructed volume converted to Cartesian coordinates.

requirements, simple structure, easy computation of the projection operator, straightforward parallelization of the projection and back-projection operations. Furthermore, the regularization schemes used in Cartesian coordinates seem to remain appropriate in cylindrical coordinates, and simple diagonal preconditioning appears to guarantee satisfactory convergence speed, even though this point should be investigated further. Therefore, the cylindrical representation can be viewed as an interesting alternative to standard, Cartesian representations.

ACKNOWLEDGMENT

Support for this work was provided by the National Sciences and Engineering Research Council of Canada (NSERC) under the Discovery Grants program.

REFERENCES

[1] B. Hamelin *et al.*, "Design of iterative ROI transmission tomography reconstruction procedures and image quality analysis," *Med. Phys.*, vol. 37, no. 9, pp. 4577–4589, Sep. 2010.

[2] J. G. Brankov, Y. Yang, and M. N. Wernick, "Tomographic image reconstruction based on a content-adaptive mesh model," *IEEE Trans. Medical Imaging*, vol. 23, no. 2, pp. 202–212, 2004.

[3] Y. Hu *et al.*, "ECG gated tomographic reconstruction for 3-D rotational coronary angiography," in *Proc. IEEE EMB Intern. Conf.*, Buenos Aires, Argentina, Sep. 2010, pp. 3614–3617.

[4] C. Mora and M. Rafecas, "Polar pixels for high resolution small animal PET," in *IEEE Nucl. Sci. Symp.*, vol. 5. San Diego, CA: IEEE, Oct. 2006, pp. 2812–2817.

[5] M.-J. Rodriguez-Alvarez *et al.*, "Exploiting symmetries for weight matrix design in CT imaging," *Math. Comput. Model.*, vol. 54, no. 7-8, pp. 1655–1664, 2011.

[6] C. Thibaudeau *et al.*, "Cylindrical and Spherical Ray-Tracing for CT Iterative Reconstruction," *IEEE Nucl. Sci. Symp.*, no. 23, pp. 4378–4381, Oct. 2011.

[7] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. SP-41, no. 2, pp. 534–548, Feb. 1993.

[8] J. Koster, "Parallel templates for numerical linear algebra, a high-performance computation library," Master's thesis, Utrecht University, The Netherlands, Jul. 2002.

Blood vessel structure reconstruction based on SART-TVM and vessel enhancement technique

Yining Zhu, Mengliu Zhao, Defeng Chen and Hongwei Li*

***Abstract*—Cardiovascular disease is one of the main threats to the human beings' health. To diagnosing the disease, a conventional method in clinic is that imaging the blood vessel with X-ray CT and then enhancing and segmenting with image processing. However, these steps are independent between each other, hence, low quality of the CT imaging results would make the flowing enhancing or segmenting not to be satisfied. In this paper, we propose a hybrid approach which involves iteration reconstruction algorithm and blood vessel enhancement technique e.g. Frangi's vesselness. This approach utilize the Frangi's vesselness to feedback the SART process in each iteration and reconstruct the structure of the blood vessel from the projection data directly in final. Experiments show that our approach has a better performance than the conventional method especially in sparse-view and low dose case.**

***Keywords*—SART-TVM; Frangi's vessel; blood vessel reconstruction.**

I. Introduction

Cardiovascular diseases were ranked as the top fatal diseases (non-communicable) to middle-aged and older adults worldwidely, and will probably cause 25 million of deaths in the year 2030 [1]. To help diagnosing the disease in clinic, varied non-invasive imaging methods are wildly applied, such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI). Medical images constructed from these methods usually contain not only blood vessel but also other structures, e.g. soft tissue,muscle and bones. Meanwhile, the density of the vessel and the blood is lower than the surrounding tissue, for which the vessel is hard to be distinguished in the images. For this case, in X-ray CT, the conventional processing method includes three steps: first, injecting contrast medium into the patient's vein; then, scanning the patient while the contrast medium is transmitted to the target organ by the blood flow and reconstructing the corresponding CT images; finally, processing the CT images to enhance the blood vessel structure, for instance, vessel enhancement or segmentation techniques. The contrast medium could make the vessel and the blood have a much higher CT value in the reconstructed images than usual, for which would be helpful to the vessel enhancement. The enhancement (or segmentation) results of the CT images could be used to be analyzed by the doctor, or for further processing, such as registration or measurement. However, it is noticed that the reconstruction of the CT image and the enhancement of the blood vessel are independent steps in usual. In this case, the low quality reconstruction images will reduce the performance of the enhancement of the vessel. Hence, in this paper, we proposed a hybrid approach which involves SART-TVM (Simultaneous Algebraic Reconstruction Technique - Total Variation Minimum) reconstruction algorithm [2] and vessel enhancement technique [3], [4]. By our approach,the structure of the blood vessel could be reconstructed directly from the projection data rather than two separate steps. Furthermore, considering applications which have high temporal resolution such as interventional therapy, we also study the blood vessel structure reconstruction from sparse-view and low dose data. Compares with the conventional method, our approach has a better effect for reconstructing the vessel structure especially in sparse-view and low dose case.

Yining Zhu is with School of Mathematical Sciences, Peking University, Beijing, 100871, China ; email: yining.zhu@pku.edu.cn.

Mengliu Zhao is with the Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, BC, Canada; email: mengliuz@sfu.ca.

Defeng Chen and Hongwei Li are with the CT Laboratory, School of Mathematics, Capital Normal University, Beijing, 100048, China.

II. Method

A. Preliminaries

1) SART-TVM algorithm: The X-ray CT imaging process usually can be written as a linear system in ideal condition,

$$\mathcal{P} = \mathbf{R}\mathcal{F}, \tag{1}$$

where $\mathcal{P}$ denotes the projection data vector, $\mathcal{F}$ is the image factor and $\mathbf{R}$ is the projection matrix. The formula (1) means that an object is expressed by expansion coefficients $\mathbf{R}$ and

generates a set of line-integration values $\mathcal{P}$ using the projection matrix. Since $\mathbf{R}$ is a huge and sparse matrix, the linear equations set is usually solved by iteration methods. SART algorithm is one of the famous method to solve the formula (1) by iteration as follows,

$$f_j^{(k+1)} = f_j^{(k)} + \frac{1}{\sum_{n=1}^{N} \mathbf{r}_{n,j}} \sum_{n=1}^{N} \mathbf{r}_{n,j} \frac{\mathcal{P}_n - \mathbf{R}_n \mathcal{F}^k}{\sum_{j=1}^{J} \mathbf{r}_{n,j}}, \quad (2)$$

where indicates f_j the j-th component (pixel) of the vector (image) $\mathcal{F}$, k is the iteration index, $\mathbf{r}_{n,j}$ is the element of $\mathbf{R}$ which means the intersection-length between the n-th X-ray and the j-th pixel in $\mathcal{F}$.

In recent years, TVM based on compressive sensing theory is a very popular research filed in CT reconstruction [5]. If $\mathcal{F}$ is a sparse (or piece-wise constant or piece-wise polynomial) image, the CT reconstruction can be regarded as a convex optimization problem by TVM,

$$\min_{\mathcal{F}} \{\frac{1}{2}||\mathbf{R}\mathcal{F} - \mathcal{P}||^2 + ||\nabla \mathcal{F}||\} \quad (3)$$

where

$$||\nabla \mathcal{F}|| = \sum_{j=1}^{J} \sqrt{(\frac{\partial \mathcal{F}}{\partial x})^2 + (\frac{\partial \mathcal{F}}{\partial y})^2}.$$

By solving the optimization problem above, it is a typical SART-TVM algorithm by which we could obtain reconstructed images with better quality especially in low dose or spares-view condition.

2) Multiscale and Frangi's enhancement: One of the most effective vessel enhancing techniques is the Frangi's vesselness [6], which is based on characteristics of eigenvalues of Hessian matrix under multiscales. Multiscale concept in computer vision field was first introduced by Witkin [7] and later developed by Koenderink [8] and Linderberg [9], who built up the fundamental theories and developed the normalization formulae. Vessels with different widths are regarded as under different scales, which could be modeled as being convoluted with Gaussian kernels using different variances. Sato [10] analyzed the behaviors toward eigenvalues of the Hessian matrix in multiscales. Suppose $\mathcal{F}(x,y,z)$ is the volume image defined on 3D domain Ω, $G_\sigma(x,y,z)$ is Gaussian kernel with standard variance σ, then the Hessian matrix under scale σ could be written as follow:

$$H_\sigma(x,y,z) = \begin{pmatrix} \frac{\partial^2}{\partial^2 x} G_\sigma & \frac{\partial^2}{\partial x \partial y} G_\sigma & \frac{\partial^2}{\partial x \partial z} G_\sigma \\ \frac{\partial^2}{\partial x \partial y} G_\sigma & \frac{\partial^2}{\partial^2 y} G_\sigma & \frac{\partial^2}{\partial y \partial z} G_\sigma \\ \frac{\partial^2}{\partial x \partial z} G_\sigma & \frac{\partial^2 a}{\partial y \partial z} G_\sigma & \frac{\partial^2}{\partial^2 z} G_\sigma \end{pmatrix} * \mathcal{F}(x,y,z)$$

Suppose $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues of $H_\sigma(x,y,z)$ and $|\lambda_| \leq |\lambda_2| \leq |\lambda_3|$. Define:

$$\begin{cases} R_A &= |\frac{\lambda_2}{\lambda_3}| \\ R_B &= \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}} \\ R_C &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \end{cases}$$

Then Frangi's vesselness ν is as following:

$$\nu(x) = \begin{cases} 0 & \text{if } \lambda_2 \geq 0 || \lambda_3 \geq 0 \\ (1 - e^{-\frac{R_A^2}{2\alpha^2}}) e^{-\frac{R_B^2}{2\beta^2}} (1 - e^{-\frac{R_C^2}{2S^2}}) & \text{otherwise} \end{cases} \quad (4)$$

The vesselness ν is close to 0 in background areas and close to 1 in center areas of vessels, which makes it very effective for vessel enhancement.

B. SART-TVM-Frangi algorithm

As we mentioned above, if the reconstructed results of blood vessel are not satisfied then the corresponding results of segmentation are also not positive. Hence, in this paper, we consider a hybrid approach which enhancing the blood vessel in the iteration process of CT reconstruction. Our SART-TVM-Frangi algorithm can be processed as follows:

1) Initialize the image as $\mathcal{F}^{(0)}$ and the corresponding result of Frangi operator $\upsilon^{(0)}$;
2) Utilize SART algorithm and Frangi operator for iteration:

$$f_j^{(k+1)} = f_j^{(k)} + \lambda \upsilon_j^{(k)} \frac{1}{\sum_{n=1}^{N} \mathbf{r}_{n,j}} \sum_{n=1}^{N} \mathbf{r}_{n,j} \frac{\mathcal{P}_n - \mathbf{R}_n \mathcal{F}^k}{\sum_{j=1}^{J} \mathbf{r}_{n,j}}, \quad (5)$$

where indicates $\upsilon_j^{(k)}$ the j-th pixel of the image calculated with the Frangi's vesselness(formula 4) after the k-th iteration, λ is relaxation factor. In order to accelerate the iteration process, the value range of λ is [2, 10] which is much bigger than the original SART algorithm.
3) Solve TVM (based on ROF model [11]) for the iteration result in the previous step;
4) Downscale the value of the image because the iteration result in step 2) is too big to match the projection equation (formula (1)),

$$\mathcal{F}_{ds}^{(k+1)} = \alpha \mathcal{F}^{(k+1)}, \quad (6)$$

where α is the scale factor calculated as follows锟斤拷

$$\alpha = \frac{\sum_{j=1}^{J} \mathcal{P}_j}{\sum_{j=1}^{J} (\mathbf{R}\mathcal{F}_j^{(k+1)})}.$$

5) Judge the current image is satisfied or not and then finish the iteration process or turn to step 2) for the next iteration.

In order to solve TVM in step 3), Chambolle gave a fast algorithm with dual-formulation in [12] as follows:

$$p^{n+1} = \frac{p^n + \tau(\nabla(div(p^n)) - f/\lambda)}{1 + \tau|\nabla(div(p^n)) - f/\lambda|} \quad (7)$$

$$u^{n+1} = (f - \lambda div(p^{n+1)}) \quad (8)$$

where u^{n+1} denotes the image calculated for TVM after the $(n+1)$-th iteration, f the original image, $div(\cdot)$ the divergence operator and τ the pseudo time step, which should be less than $1/4$ for 2D images to guarantee convergence. Meanwhile, we also demonstrate the flow chart of the above process in Fig.1. In SART-TVM-Frangi algorithm, the pixel f_j is calculated by

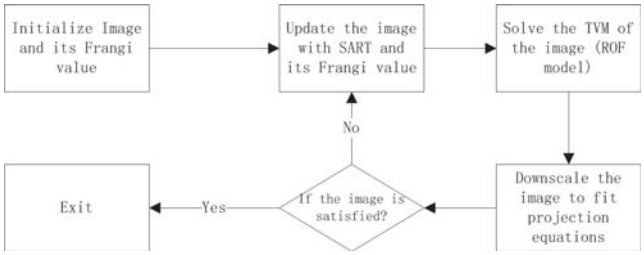


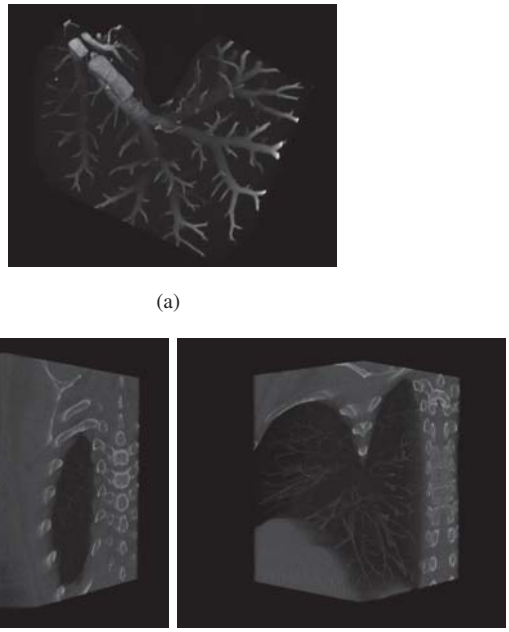
Fig. 1: The flow chart of the SART-TVM-Frangi algorithm.

Frangi operator for the possibility of whether in blood vessel in each iteration and then allocated the residual error adaptively to enhance the blood vessel and weaken the other tissues so as to reconstruct the vessel structure directly. It is should pointed out that the result reconstructed by our approach is not the real CT value but the structure information of each pixel e.g. the blood vessel or others. For practical application, in some sense, the information which presents the blood structure is more useful than the simple CT value.

III. Experiments

We test our approach with two real medical CT images as shown in Fig.2. The liver image came from the biotechnology department of Shanghai Jiaotong University and the lung image came from ISBI's VESSEL12 challenge [13]. These two CT images are reconstructed from the data scanned with full-view and ordinary dose.

We first simulate the projection data in the low dose condition (about 30% of the ordinary dose) with two CT images and then select fewer views in the two projection data sets so as to form spare-view cases.



(a) (b) (c)

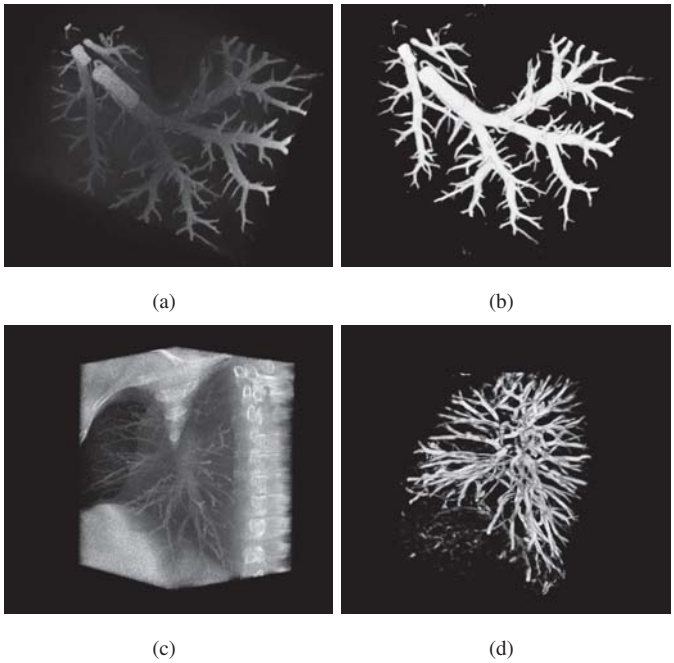
Fig. 2: The organs which contain blood vessel: (a) is the liver; (b), (c) are the lung with the view inside and outside respectively.

In the low dose case, as shown in Fig. 3 (a) and (c), the conventional reconstruction method such as SART-TVM will lead a low quality of the CT image caused by the noise and the blood vessel is difficult to distinguished especially in the lung. However, the blood vessel structures by our approach are demonstrated well. As we mentioned above, Frangi operator could enhance the vessel and weaken other structure as well as noise in the reconstruction iterations.

As shown in Fig. 4, the blood vessel structures are reconstructed with sparse-view projection data. The results by the SART-TVM algorithm are blurred with the surrounding tissue while the images reconstructed by our approach still remain blood vessel structures.

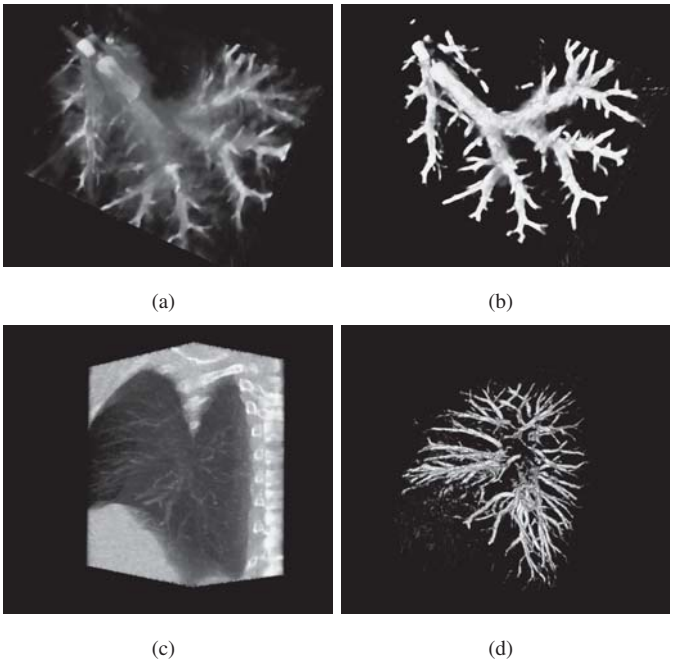
IV. Conclusion

In this paper, we proposed a hybrid approach to reconstruct the structure of blood vessel from projection data directly. This approach integrates vessel enhancement technique to the iteration of reconstruction process by using Frangi operator and derives the structures of the blood vessel. Experiments on real CT data show that our approach has a better performance than conventional reconstruction algorithm such as SART-TVM in the low dose and sparse-view case.



(a) (b) (c) (d)

Fig. 3: The reconstruction results of blood vessel in liver and lung with low dose: (a), (c) are reconstructed by SART-TVM; (b), (d) are reconstructed by our approach.



(a) (b) (c) (d)

Fig. 4: The reconstruction results of blood vessel in liver and lung with low dose and sparse-view: (a), (c) are reconstructed by SART-TVM ; (b), (d) are reconstructed by our approach. The livers ((a) and (b)) are reconstructed with 10 views while the lungs ((c) and (d)) are 30 views.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under the grants 60971131, 61127003 and 60972140, Beijing Education Committee under the grants PHR20110509 and KZ201110028034.

REFERENCES

[1] W. H. O. (WHO), "World health statisics 2012," 2012.

[2] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[3] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys (CSUR)*, vol. 36, no. 2, pp. 81–121, 2004.

[4] S. Jbabdi, P. Bellec, R. Toro, J. Daunizeau, M. Pélégrini-Issac, and H. Benali, "Accurate anisotropic fast marching for diffusion-based geodesic tractography," *Journal of Biomedical Imaging*, vol. 2008, p. 2, 2008.

[5] E. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-Ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[6] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Interventation —MICCAI' 98*, ser. Lecture Notes in Computer Science, W. Wells, A. Colchester, and S. Delp, Eds. Springer Berlin Heidelberg, 1998, vol. 1496, pp. 130–137.

[7] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, vol. 9, mar 1984, pp. 150 – 153.

[8] J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984. [Online]. Available: http://dx.doi.org/10.1007/BF00336961

[9] T. Linderberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 117–156, 1998.

[10] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol. 2, no. 2, pp. 143 – 168, 1998.

[11] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[12] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1, pp. 89–97, 2004.

[13] I. I. S. on Biomedical Imaging, "Vessel segmentation in the lung 2012," 2012. [Online]. Available: http://vessel12.grand-challenge.org/

A Novel Dual-energy CT Scanning Strategy and Its Image Restoration Method

Jia Hao, Li Zhang, Zhiqiang Chen and Kejun Kang

***Abstract*— Dual-energy CT has better ability in material discrimination and quality assurance. In this paper, we propose a novel dual-energy CT scanning strategy using two X-ray tubes. The low-energy scans the object with normal dose, while the high-energy scans the object with much sparser view to reduce the radiation dose. In this condition, the high-energy attenuation map suffers from severe artifact and noise. As we know, the low-energy and high-energy images should have the same structure as they scan the same object and the images should be well aligned. Based on this knowledge, a structure similarity based restoration method is proposed for this scanning strategy. The structure information is obtained by a non-local pixel similarity measurement in the low-energy attenuation map, and then the high-energy attenuation map is restored by weighting average with the similarity relationship established. Using the method the artifact is effectively suppressed and the image quality is improved. With this novel dual-energy CT scanning strategy and reconstruction method, the total dose is reduced and the image quality is well preserved. Experiments have been conducted to demonstrate the effective of the proposed configuration and method.**

***Index Terms*— Dual-energy CT, sparse view, image restoration, artifact reduction.**

I. Introduction

Although dual-energy CT was first conceived in the 1970s, it was not widely used for CT indications since recent years [1]. In 2005, Siemens launched their CT SOMATOM Definition, which generates two spectra by installing two x-ray tubes into one CT [2]. Recently, the simultaneous acquisition of volumetric dual-energy data has been introduced using multidetector CT (MDCT) with two X-ray tubes and rapid kVp switching (gemstone spectral imaging). Two major advantages of DECT are material decomposition by acquiring two image series with different kVp and the elimination of misregistration artifacts [3].

Dual-source CT is composed of two X-ray tubes and two detectors arranged at an angular off-set on the rotating gantry (see Fig.1). This configuration allows for the evaluation of the coronary arteries at a high temporal resolution or for the acquisition of dual-energy CT data. When both X-ray tubes are operating at different tube voltages, two different X-ray spectra are simultaneously obtained, giving rise to dual-energy tissue characterization. Due to the angular off-set of the tube-detector combinations, image registration is needed and dual-energy processing has to be performed in the image space. Since the images are dependent on the attenuation of the x-ray beam, which depends on the voltage applied across the tube, each image acquired is energy dependent. Attenuation is also dependent on the density of the material through which the beam passes, and knowing the energy of the beam allows assumptions to be made about the attenuating material based on the spectral properties of the detected radiation.

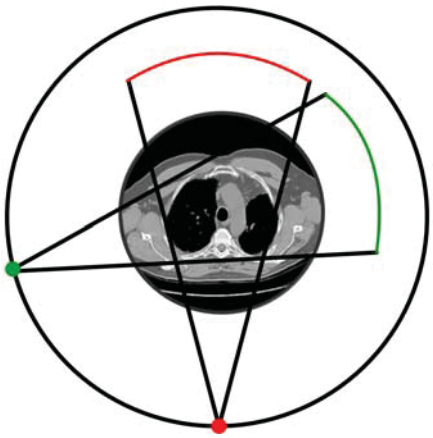


Fig.1 Dual-energy CT scanning geometry with two X-ray tubes.

With the rapid development of CT scanning, radiation dose becomes a significant concern for both patients and doctors. Different dose reduction strategies have been proposed and researched. In this paper, we propose a novel scanning strategy for dual-source CT. Reducing the sampling number during CT scanning is a common and popular method. Here we follow the similar idea and design a novel scanning configuration. The low-energy X-ray tube scans the object with normal dose while the high-energy scans the object with lower dose by reducing the view number. After image registration the low-energy and high-energy attenuation map should have the same structure and edge as they represent the same object. Then we propose a non-local image restoration scheme for this kind scanning geometry, which effectively reduce the artifact and will preserve the image quality.

This work was partly supported by the National Natural Science Foundation of China (Grant No. 10875066 and No. 10905030).

All the authors are with 1)Department of Engineering Physics, Tsinghua University, Beijing, 100084. China. 2) Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing, 100084. China

Hao Jia, PhD candidate, Email: haoj04@mails.tsinghua.edu.cn.

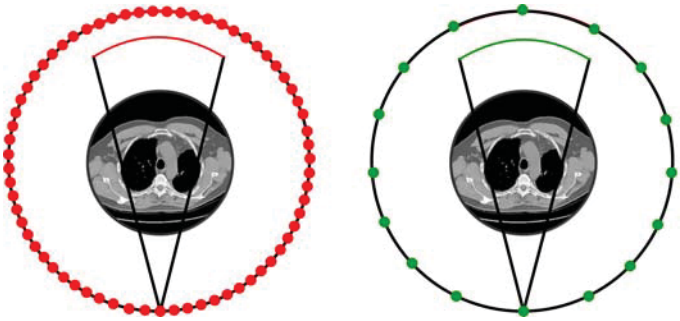
Zhang Li, Professor, Email: zhangli@nuctech.com

II. BACKGROUND AND METHODOLOGY

A. A Novel Dual-energy Scanning System

The proposed dual-energy CT configuration uses two rotating tubes to acquire both high and low voltage images. One of the tubes scans the object with normal dose and another scans with sparse view, as presented in Fig.2. The two tubes rotate simultaneously, however they works on different voltage and pulse frequency in generating X-ray.

In medical application, low-energy X-ray has better ability to discriminate low-contrast materials. Thus, the low-energy tube scans the object with normal dose. The high-energy tube scans with fewer angular views. Using this scanning geometry, the reconstructed high-energy attenuation map suffers from severe artifacts and the image quality is significantly degraded.



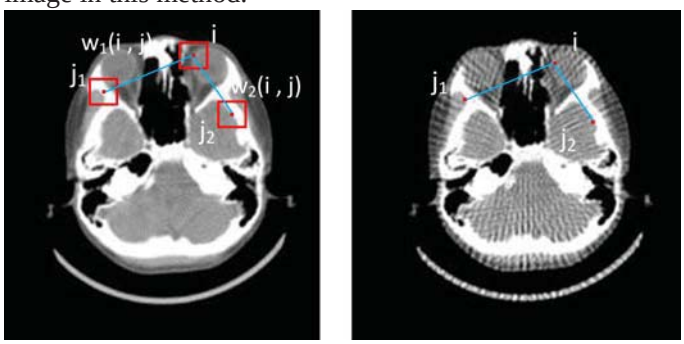
(a) (b)

Fig.2 The proposed scanning configuration. (a) and (b) represent the low and high X-ray scanning, respectively. The low-energy X-ray tube rotate around the object and generates X-ray with full angular sampling while the low-energy X-ray tube scan the object with reduced sampling.

B. Non-local Image Restoration Algorithm

Different methods have been proposed to reduce under-sampling artifacts. The most attractive and popular method is based on compressed sensing theory. In 2006, Sidky et al [4] proposed a constrained, total-variation minimization algorithm for sparse view scanning configuration. However the iterative reconstruction method needs quite long computational time. Consider that the low and high energy attenuation maps should have strict same structure and edge, as they scan the same object and are well registered. The low-energy attenuation map should have useful information for the high-energy image restoration. Thus, we establish a relationship between the two images by a non-local scheme. Structure information is first extracted from the normal dose image by pixel similarity computation. The reconstructed under-sampling image should follow the same structure similarity with the normal dose image. Thus we design a non-local image restoration method for under-sampling artifacts reduction.

Specific to the proposed dual-energy CT configuration, low-energy attenuation map is artifact-free and has much better image quality. The structure and detail information are well preserved. To restore the high-energy attenuation image, the structure information of low-energy image is helpful. In this condition we can adopt the attenuation map as a reference image and restore the effective atomic number image. The proposed method is motivated by non-local means which achieve excellent results in image noise reduction [5]. The implementation of this algorithm is shown in Fig.3, and the low-dose image serves as the reference image in this method.



(a) (b)

Fig.3 Searching scheme and weighting function establishing method. (a) is the reconstructed low-energy attenuation image with normal dose, it serves as the reference image; (b) is the under-sampling high-energy attenuation map from the conventional CT reconstruction method.

Then we briefly introduce our method. Suppose the reference image is expressed as I_{ref}, and the degraded image to be restored is expressed as I. The two images are strict registered and have the same structure. For a selected pixel i, the value is $\mu(j)$. Compute the weighting coefficients $w(i, j)$ in I_{ref} between pixel i and the other pixels in the reference image. It is measured by the distance between two $n \times n$ blocks N_i and N_j centered at pixel i and j respectively. The distance $d(i, j)$ can be calculated by a Gaussian-weighted Euclidean distance between the two blocks:

$$d(i, j) = \left\| N(j) - N(i) \right\|_{2,a}^2 \quad (1)$$

a is the standard deviation of the Gaussian function. The weighting coefficient between pixel i and j can be formulated as:

$$w(i, j) = \exp\left(-\frac{d(i, j)}{h^2}\right) = \exp\left(-\frac{\left\| N(j) - N(i) \right\|_{2,a}^2}{h^2}\right) \quad (2)$$

where h is a smooth parameter. For all the pixels in I_{ref}, compute the weighting between the selected pixel i and the other pixels. In the image to be restored, weighted all the pixels use the weighting value in the corresponding pixel in the reference image, and the pixel value after restoration $R(\mu(i))$ is obtained as:

$$R(\mu(i)) = \frac{w(i,j)\mu(j)}{\sum w(i,j)} \quad (3)$$

In this method, the weighting function $w(i,j)$ is computed in the reference image however the weighting average step is implemented in the high-energy attenuation image.

It is also worth mentioning that the statistic value accuracy is not impacted by this method, because of the normalized weighting average is only a numerical redistribution step. The image singularities and structure errors are smoothed by the weighting processing, however, the mean value in a selected region maintains the same. If the reference image is the same with the current image, it will fall back to the non-local means denoising method.

III. EXPERIMENT

Experiments are implemented on an explosive detection dual-energy CT system. The low and high X-ray spectrum used here is presented in Fig.4. In total of 360 projections are acquired during the low-energy X-ray scanning, while only 60, 30 and 15 projections are acquired using high-energy spectra. The dose from high-energy X-ray is reduced from 1/6 to 1/24. The original reconstructed images are shown in Fig.4. (a) is the low-energy attenuation map and (b) is the high-energy attenuation map. It can be found severe under-sampling artifacts, which degrade the image quality.

Using the proposed image restoration method, the under-sampling artifacts are significantly reduced. The results are shown in Fig.6. In the first line, reconstructions without processing are displayed. They are reconstructed from 60, 30 and 15 projection views respectively. The second line are the restoration images, which can be found much better in image quality and SNR. Quantitative comparison is shown in Fig.7. In dual-energy CT, reconstructed value is always important in material discrimination. Here we use a full-view reconstruction image as the ground-truth. The results demonstrate that our method has excellent performance in accurate reconstruction for under-sampling dual-energy CT reconstruction.

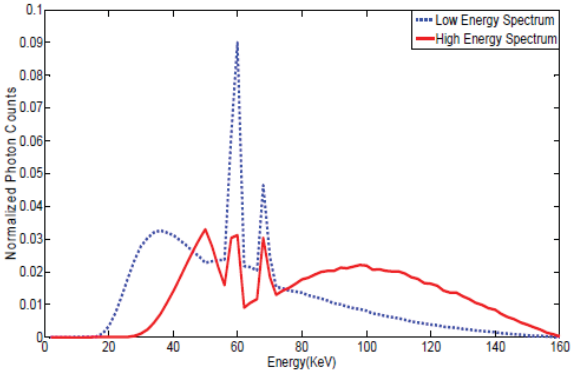
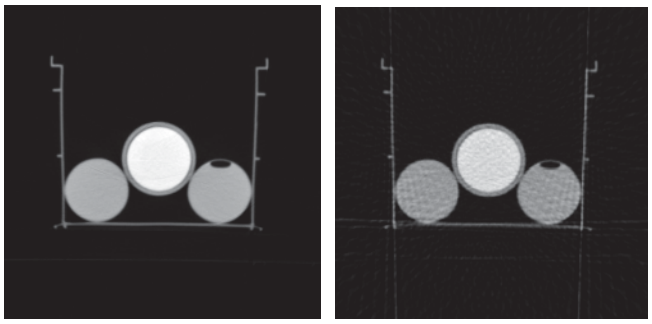
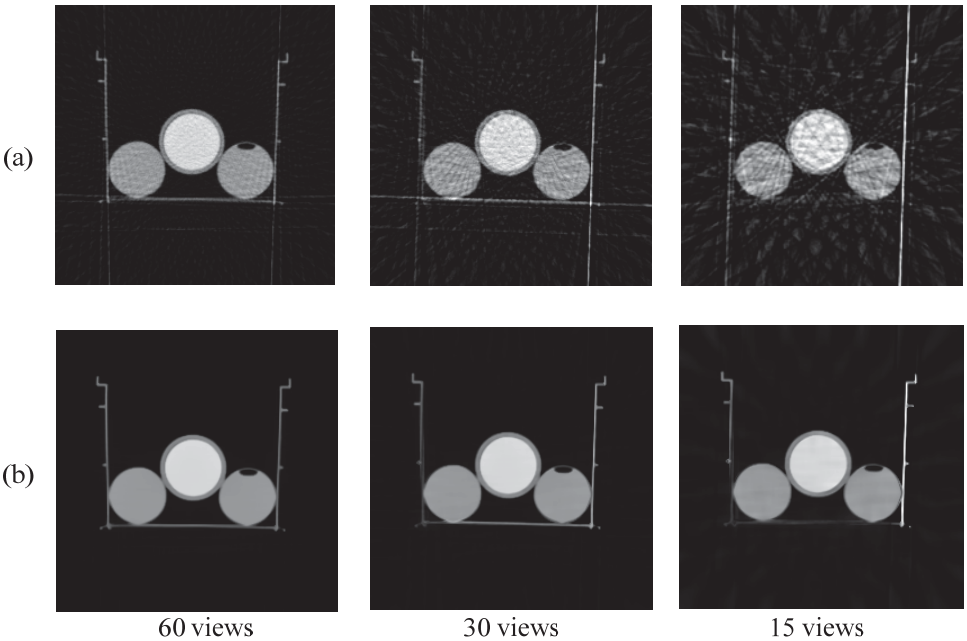


Fig.4 The high and low energy spectrum used in the experimental dual-energy CT system.



(a) (b)

Fig.5 The reconstruction from the proposed dual-energy CT scanning configuration. (a) is the low-energy attenuation image from 360 projection views and (b) is the high-energy attenuation image from 60 projection views.



(a)

(b)

60 views 30 views 15 views

Fig.6 The original and processed high-energy attenuation images from the proposed dual-energy CT scanning configuration. (a) is the original images without any processing. The total view number is 60, 30 and 15 respectively. (b) is the restoration images using the proposed non-local method. The greyscale display window is [0, 0.03].

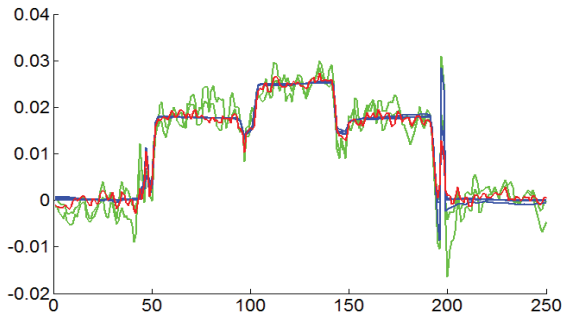
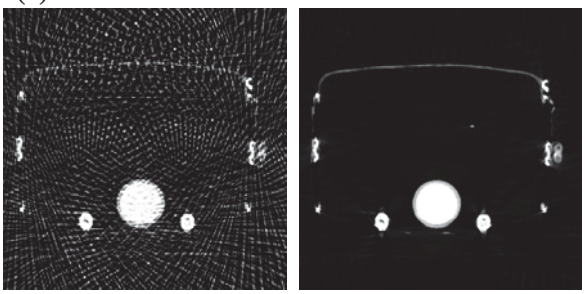


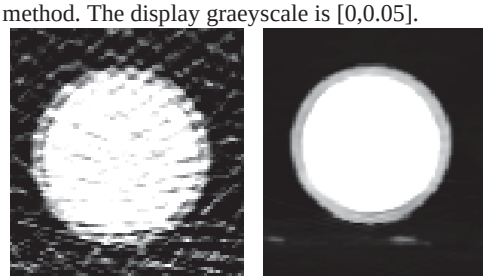
Fig.7 Pixel value comparison along the profile of the reconstructed images.The red line can be recognized as the ground-truth reconstructed from 360 projection views. The green lines show the value from under-sampling projections without processing in Fig.5(a). The blue lines indicate the restoration image results in Fig.5(b).

To demonstrate the effective of the proposed method, we use a more complicated object. A draw-bar box with a phantom was scanned in the dual-energy CT. The number of angular views of high-energy X-ray projections is 60, while the low-energy view number is 360. The reconstructed high-energy attenuation image is shown in Fig.8 (a), and the restoration is shown in Fig.8 (b).



(a) (b)

Fig.8 The high-energy attenuation maps reconstructed from a under-sampling dual-energy CT configuration. (a) is the original image from 60 views and (b) is the restoration result using the proposed method. The display graeyscale is [0,0.05].



(a) (b)

Fig.9 The zoomed in image of the images in Fig.7. (a) is the original reconstruction from the under-sampling projections, and (b) is the restoration result using the proposed method. The display graeyscale is [0,0.05].

IV. DISCUSSION

A. *Why this algorithm works?*

The proposed method establishes a relationship between the low and high energy attenuation map. The structure information is obtained from the full-sampling projection image. Thus, this method preserves the structure and the value accuracy.

It differs from previous compressed sensing based methods in that: this method is only an image restoration method. It is a post-processing method after reconstruction, and it is much faster than iterative method. The results are better than TV constrained iterative method as we have researched. Also, the strategy can be extended to the other applications: when there is a high-quality reference image, it can be used to process a noisy or artifact degraded image. The two images should have the same structure and be well aligned. Dual-energy CT is only one of the applications of this method, as it can obtain high-energy and low-energy attenuation maps, and they have strict the same structures.

B. *Algorithm optimization*

This method has been demonstrated effective in artifact and noise reduction. However, the searching scheme is quite slow. The high computational complexity is due to the cost of weights calculation for all pixels in the image. In practical application, the weighting calculation area is limited to a smaller window instead of the whole image for computational aspects. In this study, processing a 256×256 image takes about 30 seconds. Other methods can be used, such as lower the searching window, using pre-classification method and so on.

V. CONCLUSION

We propose a novel dual-energy CT scanning configuration. One of the tubes generates X-ray with lower frequency, thus the projection sampling is reduced. With the conventional CT reconstruction method, severe under-sampling artifacts degrade the image quality. We introduce a non-local restoration algorithm for this scanning configuration, which uses the high-quality image as reference information and process the under-sampling reconstruction image. Experiments demonstrate the effective of this method.

This is only a preliminary research and result. There are a lot of problems to be further studied. As we know, attenuation maps under low and high energy X-ray may differ significantly, especially near the K-edge of the scanned material. In this condition, the proposed algorithm should be carefully evaluated. And here the scanned object is a little simple. We will further our study with different phantoms and samples, in order to demonstrate the configuration and method.

REFERENCES

[1] R.E. Alvarez and A. Macovski, “Energy selective reconstructions in x-ray computerized tomography,” Phys. Med. Biol., vol. 21, no. 5, pp. 733–744, 1976.

[2] T. G. Flohr, C. H. McCollough, H. Bruder, et al. “First performance evaluation of a dual-source CT (DSCT) system” Eur Radiol, vol. 16, pp. 256–268, 2008.

[3] Z. R.Ying, R. Naidu, C. R. Crawford. “Dual energy computed tomography for explosive detection”, Journal of X-ray Science and Technology, vol. 14, pp. 235-256, 2006.

[4] E.Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," Journal of X-ray Science and Technology, vol. 14, pp. 119-139, 2006.

[5] A. Buades, B. Coll, J. M. Morel, “A review of image denoising algorithm, with a new one.” Multiscale Model. Simul. vol. 4, no. 2, pp. 490-530, 2005.

Top-Level Design of the First CT-MRI Scanner

Ge Wang, *Fellow, IEEE*, Feng Liu, Fenglin Liu, Guohua Cao, Hao Gao, Michael W. Vannier

***Abstract*—Omni-tomography is conceptualized based on interior tomography developed over the past five years. By omni-tomography, we envision that the next stage of biomedical imaging will be the grand fusion of many tomographic modalities into a single gantry ("*all in one*") for simultaneous data acquisition of many complementary features ("*all at once*"). This integration has great potential, because physiological processes are often dynamic and complicated, and must be observed comprehensively and promptly. As an inspiring simple example of omni-tomography, here we design the first CT-MRI scanner for vulnerable plaque characterization, and suggest exciting research opportunities.**

***Index Terms*—Compressive sensing, interior tomography, modality fusion, omni-tomography, CT-MRI, vulnerable plaque characterization**

I. Introduction

While CT, MRI, PET, SPECT, and ultrasound have their well-defined roles for medical imaging, over the past decade multi-modality systems become increasingly popular, such as PET-CT and PET-MR scanners for sequential or contemporaneous data acquisition. However, these paired modalities impose limitations that may compromise our understanding of physiological processes relative to fine details and rapid changes driven by a beating heart, and not all pairs of tomographic scanners can be engineered closely together or within a single gantry, given the bulkiness of individual systems such as CT and MRI scanners and the conflict in imaging physics such as rotating metallic parts for CT and magnetic fields for MRI.

Recently, our group has been performing pilot studies on what is called omni-tomography [1]. This strategy is conceptualized based on interior tomography developed over the past five years [2]. By omni-tomography, we envision that the next stage of biomedical imaging will be the grand fusion of many tomographic modalities into a single gantry ("*all in one*") for simultaneous data acquisition of many complementary features ("*all at once*").

Now, novel multi-functional probes, multi-physics modeling, high-tech engineering and advanced image reconstruction, especially interior tomography and compressive sensing, present new opportunities to peek into living biological systems non-invasively, systematically and concurrently. Building blocks are either available or emerging for the initial development of omni-tomography. As an inspiring case of omni-tomography, here we design the first CT-MRI scanner for vulnerable plaque characterization, described as the "*holy grail*" of cardiology in news such as http://www.nytimes.com/2006/11/27/business/worldbusiness/27iht-heart.3681162.html?_r=0.

With the CT-MRI scanner, CT and MRI scans can be seamlessly merged for spatiotemporal registration, extendable to include other imaging modalities. CT provides structural definition in snapshot. MRI reveals blood flow, soft tissue contrast, functional, cellular and molecular features. Neither CT nor MRI itself could cover all biomarkers of vulnerable plaques, including cap thickness, lipid-core size, stenosis, calcification, hemorrhage, elasticity, inflammation, endothelial status, oxidative stress, platelet aggregation, fibrin deposition, enzyme activity, microbial antigens, apoptosis and angiogenesis, and so on. It would be exciting to have all of these features in high spatial, contrast, temporal resolution within a common coordinate system. Even if MRI of vulnerable plaques could have sufficient resolution and speed, retrospective image registration between CT and MRI is not a desirable alternative, because of registration errors due to non-repeatable contrast dynamics, organ motion and deformation, MRI-induced geometric distortion and signal nonlinearity, as well inconsistent contrast mechanisms between CT and MRI.

In this paper, we present a top-level engineering design of the first CT-MRI scanner, which is a major refinement of our initial concept outlined in [1]. The key idea is to let CT and MRI focus on a relatively small region of interest (ROI). Hence, each imaging modality can be simplified and parallelized. Then, they can be accommodated within a single gantry without physical interference. In the next section, we

This work was supported in part by NIH/NIBIB (R01 EB011785), NSF/MRI (0923297), and an imaging fund from Rensselaer Polytechnic Institute, Troy, New York, USA.

G. Wang is with the Biomedical Imaging Center, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, USA (ge-wang@ieee.org)

F. Liu is with School of Information Technology & Electrical Engineering, University of Queensland, Brisbane, Qld. 4072, Australia (feng@itee.uq.edu.au)

F. L. Liu is with the Key Lab of Optoelectronic Technology and System, Ministry of Education, Chongqing University, Chongqing 400044, China (liufl@cqu.edu.cn), and currently a visiting scholar with the Biomedical Imaging Center, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, USA

G. H. Cao is with School of Biomedical Engineering & Sciences, Virginia Tech, Blacksburg, VA 24061, USA (ghcao@vt.edu)

H. Gao is with Departments of Mathematics and Computer Science, and Radiology and Imaging Sciences, Emory University, Atlanta, GA 30322, USA (hao.gao@emory.edu)

M. W. Vannier is with Department of Radiology, University of Chicago, Chicago, IL 60637, USA (mwvannier@gmail.com)

GW: Overall; FL: MRI subsystem; FLL, CHC: CT subsystem; HG: reconstruction; MWV: biomedical relevance

first present our overall design, and address key points for the CT and MRI sub-systems respectively. In the third section, we show some initial simulation results. In the last section, we discuss relevant issues and conclude the paper.

II. Top-Level Design

Traditional CT methods cannot exactly reconstruct an interior ROI solely from truncated projections along x-rays through the ROI. This interior problem has been studied for decades, and the fact that precise reconstruction could not be obtained from local data contributed to the long-standing CT architectures whereby detectors always fully cover a transverse slice. Now, interior tomography allows for theoretically exact reconstruction over an ROI from purely local data aided by practical prior knowledge. More importantly, interior tomography has been extended as a general imaging principle and developed for other modalities such as MRI [1], SPECT [3], and others. Thanks to interior tomography, data acquisition module of different types can be made more compact or sparse to provide space and reduce cost for omni-tomography.

A. *Overall Description*

Our proposed CT-MRI system is illustrated in Figure 1. This design is mainly based on the generalized interior tomography.

B. *CT Sub-system*

As shown in Figure 2, the CT sub-system consists of 9 x-ray tubes, each of which is paired with a detector array. This allows instantaneous data acquisition of 9 views targeting a heart. Nine views could be enough in special cases but are generally insufficient for ROI reconstruction of diagnostic quality. Hence, the multi-source data acquisition assembly can be rotated for 3 or 5 sets of 9 views. Each data acquisition session can be synchronized to the ECG signal with breath-holding. In an alternative design, 45 or more carbon-nano-tube (CNT) x-ray focal spots can be distributed along half a circle or a full circle for rapidly multiplexed acquisition in 9-view groups.

The x-ray tubes in the proposed CT-MRI system will be under a significant magnetic field (about 0.5T with the current design for a 1.5T background field in the field of view). Hence, the electron beams in the x-ray tubes will interact with the magnetic field. The magnetic field can affect the trajectories of both the primary electrons (i.e., the electron from the cathode) and the backscattered electrons (i.e., the electrons backscattered from the anode). The primary electrons can be deflected by the magnetic field [4]. This deflection problem is more severe when the magnetic field is oblique to the cathode-anode electric field in the x-ray tube. Even if the magnetic field is parallel to the cathode-anode electric field, electron beam defocusing remains problematic, because electrons emitted from a thermionic cathode are omnidirectional and have a distribution of initial velocities. When the accelerated primary electrons strike the anode, a significant portion of those electrons will be backscattered. For an x-ray tube working at 65kV, more than 50% of the electrons incident on the target are backscattered [5]. Without a magnetic field, only a small fraction of these backscattered electrons will return to the anode. However, when a strong magnetic field exists, the backscattered electrons will have an elevated probability of striking the anode again, resulting in an increased total x-ray output, slightly softening the x-ray spectrum and enlarging the focal spot size [5].

In the proposed CT-MRI system, the x-ray tubes will work in the static main magnetic field from the main magnets without complication of the magnetic fields from the gradient and RF coils. The main magnetic field is parallel to the longitudinal direction of the CT-MRI system. Using a reflection-type anode at a suitable take-off angle, the cathode-anode axes in all the

principle and two pieces of prior art. First, an existing open MRI architecture gives open space between two donut-shaped magnets for CT. Second, the work on combination of x-ray imaging and MRI shows the compatibility of x-ray imaging and MRI [4].

Figure 1. Rendering of the proposed CT-MRI scanner which is a special case of omni-tomography.

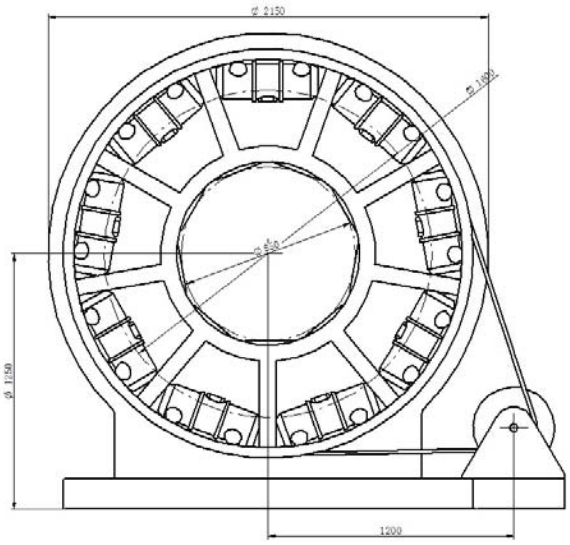


Figure 2. Geometric parameters of the proposed CT sub-system utilizing 9 x-ray tubes that can be rotated and physiologically gated if needed.

x-ray tubes can be made parallel to the direction of the magnetic field. Our simulation study indicates that a 0.5T static magnetic field is expected at the x-ray tube position. To prevent the backscattered electrons from returning to the focal spot area, active shielding can be used to minimize the magnetic field near the focal spot [6].

With our design, the state of the art CT image quality can be duplicated except for improved temporal resolution. When sufficient prior information is available, instantaneous 9-view reconstruction can reveal features of interest in special cases. If more data are necessary for more general applications, the 9-view-based acquisition process can be repeated at different angular positions by rotating conventional sources with physiological gating or multiplexing novel CNT tubes nearly in real-time. Spatial resolution can be initially set to 0.3-1mm, which can be refined as needed. The relationship between image noise and radiation dose is well known, and will be used to optimize imaging protocols.

C. *MRI Sub-system*

The MRI sub-system uses a highly sophisticated superconducting electromagnet technology. The performance of the paired superconducting magnets is critical to image quality. The static field produced by the magnets in the field of view is relaxed to be much smaller than the counterpart for a conventional MRI scanner, and needs to be strong and nearly perfect (a few parts-per-million (ppm) variation) only throughout a cardiac ROI. This represents a major simplification relative to the requirements for a much larger conventional field of view. The design of the split architecture with a large central gap (more than 40cm) is challenging, demanding the full consideration of the electromagnetic, mechanical and thermal properties of all the involved components.

As shown in Figure 3, our superconducting MRI magnet has a two-layer coil configuration. The first layer with a small radius (50cm) provides a primary magnetic field in the field of view, while the second layer with a large radius (80cm) is mainly for shielding the stray magnetic field within the domain of interest (8m in length and 5m in radius). The magnet design offers a 1.5T field of view of an appropriate diameter (25cm), which defines a spherical field of view. The field uniformity inside the spherical field of view is 1ppm (peak-peak). This magnet design allows the integration of CT in the gap between the pair of magnets. During the design, a current density pattern was first derived with a simplified linear field-source relationship. Then, wire layouts were determined using a nonlinear optimization scheme [7]. A quench simulation study was conducted to balance imaging performance and system cost. Since engineering errors unavoidably compromise field uniformity, a shimming technique must be used for field correction. In this project, a robust field-based passive-shimming algorithm [8] was used to control the field uniformity and harmonics. Moreover, current-based active shimming techniques were employed to improve the field homogeneity throughout the spherical field of view.

Three orthogonal gradient coils were designed for the split magnet bore to offer linear gradient fields. In particular, paired saddle coils were used for the x- and y- gradient coils. A Maxwell pair served as the z- gradient coil. More sophisticated

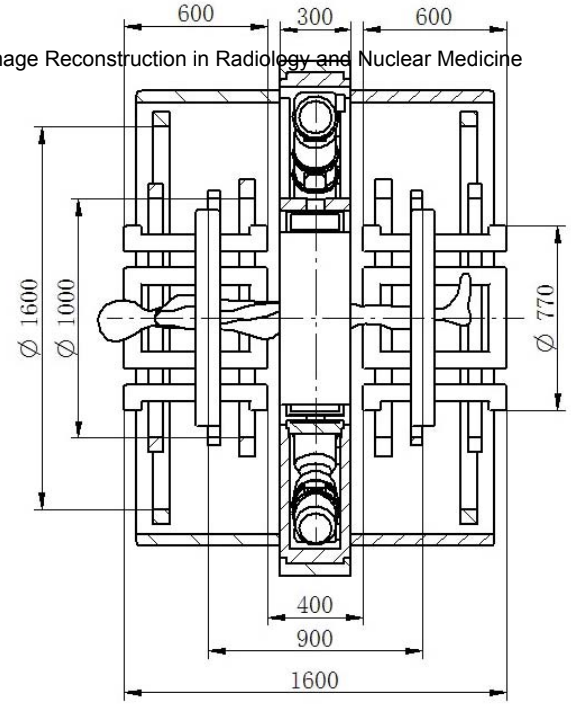


Figure 3. Geometric parameters of the proposed MRI sub-system with a split magnetic core and a two-layer coil configuration.

coil patterns are possible for advanced MRI applications [9]. The eddy current effect can be minimized through coil refinement, shield optimization, and gradient pre-emphasis. RF coils can be constructed with aluminum to be transparent to x-rays. To transmit and acquire RF signals, array coils can be arranged around the spherical field of view [10]. Also, the interaction between CT components and MRI coils can be minimized with passive and active shielding.

With our design, the current 1.5T MRI performance can be achieved. Temporal resolution can be initially set to 20-60ms or less. In-plane spatial resolution can be specified at 1-2mm, which is adequate for most cardiac function studies, although higher spatial resolution is possible when CT-based motion correction techniques are developed. SNR depends on multiple factors such as field strength, coil, sequence, motion correction, image reconstruction, and so on.

III. INTERIOR IMAGE RECONSTRUCTION

We performed a preliminary feasibility study on interior CT reconstruction [11] with 27 projections. The simulated data were used. Typical images are in Figure 4, which were reconstructed using the filtered backprojection (FBP) algorithm and an iterative algorithm based on the tensor framelet (TF) [12, 13].

The numerical phantom is a 512×512 cardiac image with a central 200×200 ROI to be reconstructed, and a coarse-resolution version of the image outside of the ROI was assumed *a priori* to facilitate the 27-view interior reconstruction. This new type of knowledge is feasible since structures outside the cardiac ROI are relatively much more stationary, and can be scanned in a low-dose setting in advance or obtained via elastic matching to a patient atlas. As shown in Figure 4, 27 projection views seem sufficient for the interior image reconstruction. In our reconstruction, a fast and highly parallel algorithm was used for computing the x-ray transform [14, 15]. With the GPU-based parallel reconstruction implemented on a notebook computer with Nvidia GTX 680M,

the TF-based reconstruction took less than 8s. The speed improvement is possible with a better algorithmic design and more advanced computing techniques.

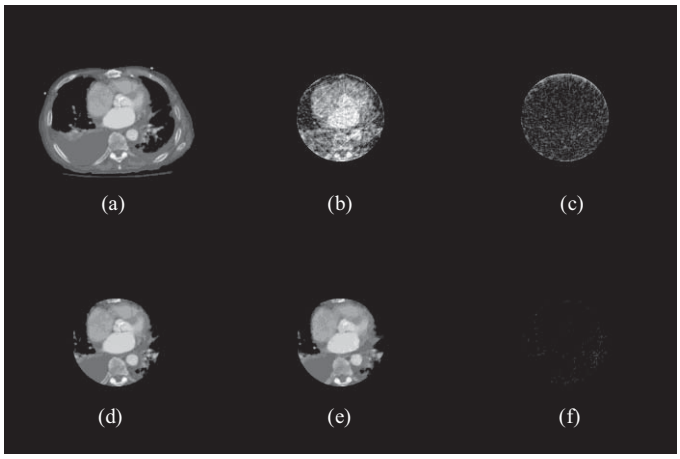


Figure 4. Interior reconstruction with the tensor framelet (TF) from 27 projection views. (a) The image phantom with a central 200×200 ROI and a coarse-resolution version of the image outside of the ROI, (b) and (e) the iimage reconstructed using FBP and TF respectively, (d) the true ROI image, (c) and (f) the error images associated with (b) and (e) respectively.

The above simulation is only the first part of our simulation design. The second part is compressive sensing based cardiac MRI reconstruction. Then, the third part is fused CT-MRI reconstruction, in hope that the synergy between CT and MRI data can be utilized to reduce the number of X-ray projections to 9 so that there would be no need for any CT scan. Currently, we are actively working on these topics and will report later if any promising results.

IV. Discussions and Conclusion

Although the above CT-MRI system design is based on physical and engineering principles, construction of a real system is not an easy task since it is significantly more complicated than any multi-modality system on the market. We could start with proof-of-concept prototypes of to show unique utilities, and then gradually move to product-ready platforms. For our ultimate goal – vulnerable plaque characterization, we would likely need 50μm spatial resolution, 20ms temporal resolution, and other cutting-edge quality indices. These are far from immediately reachable but do seem unprecedented research opportunities.

While the multi-modality reconstruction is currently performed modality-wise, we hypothesize that omni-tomographic reconstruction can be done in a dictionary/atlas based framework. The most detailed domain knowledge could be put into a composite dictionary/parameterized atlas in the involved imaging mechanisms. The unified iterative reconstruction could go as follows: (1) based on the atlas an initial composite image is set; (2) an image is updated with an omni-tomographic scan; (3) a diffeomorphism is driven via flow (such as Ricci flow used to solve the Poincaré conjecture) by the dictionary/atlas, yielding a Beltrami coefficient (BC) distribution; (4) the current image is refined to minimize the data discrepancy, BC-based sparsity and other regularizing terms including cross-modality similarity. A scheme is to combine all the fidelity and penalty terms into one objective function, and minimizes it using a split-Bregman-type scheme.

Omni-tomography offers biological, technical, physical, mathematical, and economic opportunities. Biologically, the "*all-in-one*" and "*all-at-once*" imaging power allows observation of well-registered spatiotemporal features *in vivo*. Physically, multi-physics modeling suggests new imaging modes for synergistic information (such as photoacoustic imaging which combines ultrasound resolution and optical contrast). Technically, a paradigm shift of system engineering is required to marry different types of imaging components. Economically, a "*one-stop-shop*" for diagnosis and intervention may be realized that could be often more cost-effective than a full-fledged imaging center with independent modalities. Omni-tomography does have limitations due to an ROI-oriented restriction, increased complexity and possible tradeoffs as more imaging contrast mechanisms are involved. Nevertheless, these are tractable with innovative technology and methods.

References

[1] G. Wang, J. Zhang, H. Gao, V. Weir, H. Y. Yu, W. X. Cong, X. C. Xu, H. O. Shen, J. Bennett, M. Furth, Y. Wang, M. W. Vannier, "Towards omni-tomography—Grand fusion of multiple modalities for simultaneous interior tomography," PLoS ONE 7(6):e39700. doi:10.1371/journal.pone.0039700, 2012 (arXiv:1106.2124v1 [physics.med-ph], June 12, 2011)

[2] G. Wang and H. Y. Yu, "The meaning of interior tomography," Phys. Med. Biol., invited topical review, to appear, 2013

[3] J. S. Yang, H. Y. Yu, M. Jiang, G. Wang, "High order total variation minimization for interior SPECT," Inverse Problems 28:015001(24pp), 2012

[4] Z. Wen, R. Fahrig, S. Conolly, N. J. Pelc, "Investigation of electron trajectories of an x-ray tube in magnetic fields of MR scanners," Med Phys 34:2048-58, 2007

[5] Z. Wen., N. J. Pelc, W. R. Nelson, R. Fahrig, "Study of increased radiation when an x-ray tube is placed in a strong magnetic field," Medical Physics 34:408-419, 2007

[6] P. Lillaney, M. Shin, S. M. Conolly, R. Fahrig, "Magnetostatic focal spot correction for x-ray tubes operating in strong magnetic fields using iterative optimization," Med Phys, 39:5567-5583, 2012

[7] S. Crozier, F. Liu, "Shielded, asymmetric magnets for use in magnetic resonance imaging, United States Patent 20060255805

[8] F. Liu, J. Zhu, L. Xia, S. Crozier, "A hybrid field-harmonics approach for passive shimming design in MRI," IEEE Trans. on Applied Superconductivity 21:60-67, 2011

[9] M. Zhu, L. Xia, F. Liu, J. Zhu, S. Crozier, "Finite difference method for the design of gradient coils in MRI – An initial framework," IEEE Trans. on Biomedical Engineering 59:2412-2421, 2012

[10] M. Hong, Y. Yu, H. Wang, F. Liu, S. Crozier, "Compressed sensing MRI with singular value decomposition-based sparsity basis," Phys. in Med. Biol. 56:6311-6325, 2011

[11] H. Y. Yu and G. Wang, "Compressed sensing based interior tomography," Phys. Med. Biol. **54**:2791-2805, 2009

[12] H. Gao, R. Li, Y. Lin, and L. Xing, "4D cone beam CT via spatiotemporal tensor framelet," Med. Phys. **39**:6943-6946, 2012

[13] https://sites.google.com/site/spatiotemporaltensorframelet

[14] H. Gao, "Fast parallel algorithms for the x-ray transform and its adjoint," Med. Phys. **39**: 7110-7120, 2012

[15] https://sites.google.com/site/fastxraytransform

Extended Axial Field of View in Radiotherapy Cone-Beam CT with Iterative Reconstruction

Andrew M. Davis[1,2], Erik A. Pearson[1,2], Charles A. Pelizzari[1], and Xiaochuan Pan[1,2]

***Abstract*—The size of cone-beam CT (CBCT) detectors in radiotherapy on-board imaging (OBI) systems limits the axial field of view (FOV). By acquiring two circular scans at different axial positions, this axial coverage can be increased. However, with analytic reconstruction, this results in cone-angle artifacts severely degrading the image in the region between the circular scans. Using iterative reconstruction techniques, we can extend the axial field of view while obtaining a better quality image in the region between these two scans.**

I. Introduction

Despite major advances in iterative reconstruction methods, analytic algorithms are still the predominant means of reconstruction in clinical CT scanners[1]. An analytic reconstruction algorithm is based on finding the analytic inverse to a continuous-to-continuous (CC) imaging model. This method of solving the inverse problem in CT reconstruction has been very successful as attested to by the prevalence of filtered backprojection (FBP) in the CT industry. FBP is implemented for cone-beam geometry via the FDK analytic reconstruction algorithm[2]. However, due to the inherent discrete nature of imaging detectors and the numerical arrays used to store images, the implementation of such a CC model is actually a discrete approximation of this model. Thus, there are strict requirements on data sampling which necessitate the use of densely sampled detectors and angular views.

In contrast to CC analytic methods, iterative reconstruction methods directly address the inherent discretization of the imaging modality by modeling it as a discrete-to-discrete (DD) imaging system. This is often expressed as a linear system of the form

$$\widetilde{g} = M\vec{f}, \tag{1}$$

where $\widetilde{g}$ is the measured data, M is the system matrix, and $\vec{f}$ is the image we wish to recover. In general, this linear system cannot be directly inverted to obtain the image. As such, iterative optimization techniques are used to solve this problem. Using a well understood implementation of this optimization problem, we show an example of how the more robust nature of iterative reconstructions methods allows for more flexibility in CT imaging.

The use of cone-beam CT (CBCT) is becoming more prevalent in clinical settings. We are particularly interested in the use of CBCT for image-guided radiotherapy (IGRT) in linac-mounted imaging systems. These typically consist of a flat-panel detector opposite a kV x-ray source to acquire circular scans which are then reconstructed with the analytic FDK algorithm. Unfortunately, the use of such a CC imaging model for reconstruction results in increasingly severe degradation of the image quality in image slices away from the central plane of the source with what are known as cone-angle artifacts.

In addition to this image degradation at the ends of the axial field of view (FOV), the axial coverage is limited by the detector size. For example, for a typical detector length of 30 cm parallel to the axial direction and a magnification factor of 1.5, the resulting axial coverage is 20 cm. Combining image degradation at the ends of the FOV with an already limited axial coverage can reduce the clinical utility of the linac-mounted imaging system.

It would seem reasonable to assume that the axial FOV could be extended by acquiring two circular scans at two different axial locations and then combining them to create an extended image. However, in the region between the two scans, the degradation from the cone-angle artifacts is the most pronounced. Thus the central region of the resultant image, which is presumably that of the greatest interest, is most afflicted by these cone-angle artifacts. With iterative reconstruction methods, it is possible to define the system matrix such that the data from both circular scans can be used to reconstruct a single image with an extended axial FOV.

To determine the extent to which iterative reconstruction methods would allow us to extend the axial FOV, we simulated two circular CT scans with different axial spacing between the two scans. Pearson et al. have previously shown that in such an acquisition scheme, iterative reconstruction methods can extend the axial FOV more effectively than FDK[3]. Here, we wanted to investigate what how much FOV coverage can be achieved with iterative reconstruction techniques.

II. Methods and Materials

We used a typical CBCT linac-mounted imaging system geometry for our simulation as this extended axial FOV approach has the potential to be clinically useful in IGRT. Although iterative algorithms allow for non-circular scanning trajectories, the two circle scan used in this study allows for direct comparison to analytic reconstruction with FDK. Also, such an acquisition trajectory can easily be acquired with current linac-mounted imaging systems.

Department of Radiation and Cellular Oncology (1), and the Department of Radiology (2), The University of Chicago, 5841 S. Maryland Avenue, Chicago IL, 60637 Email: amdavis@uchicago.edu

In our simulations, each circular scan projected 330 uniformly spaced views over 360° onto a 40 by 30 cm detector with a 1.5 magnification factor. With this geometry, the maximum spacing of the two circular scans is 20 cm offset. At this maximum spacing, the 20 cm axial coverage that can be achieved with one detector is doubled. There is only a small amount of data redundancy between the data acquired from each circular scan from a slight overlap of the 30 cm detector along the central plane at each scan position. In these simulations we simulated dual circular scans spaced 10 and 20 cm apart.

The first phantom we used was a Defrise-style geometric disk phantom consisting of a 15.24 cm outer diameter tube inside of which are 5 mm thick alternating high and low density disks[4]. We produced the forward projections analytically[5]. For the disk phantom, we also produced a voxelized truth image in addition to projections at scan spacings of 12, 14, 16, and 18 cm for a comparison of the reconstructions to the truth as a function of spacing.

The other phantom we used was the anthropomorphic XCAT phantom from which forward projections were also produced analytically[6]. In the case of the XCAT phantom, the maximum 20 cm circular spacing extended the FOV into the upper torso. To avoid truncation artifacts, we doubled width and the number of bins of the detector in this proof-of-concept study.

The forward projection used for both of these phantoms attempts to model the physical projections created by the incident beam. For both phantoms, we used a polychromatic spectrum based on Monte Carlo simulations of our current Varian On-Board Imager (OBI) systems. The simulated projections are noise free.

We chose the standard MLEM algorithm for the iterative reconstruction as its use is well established. For the disk phantom, we reconstructed onto a 256x256 image array in the transverse plane using 1 mm isotropic voxels to determine the axial length. For the XCAT phantom, we reconstructed onto a 256x256 image array in the transverse plane using 2 mm isotropic voxels.

As a comparison with FDK, we reconstructed each circular scan separately using a standard Hann filter. To combine the two separate reconstructed images, we used the central plane between the two circular scan planes as a delimiter. Slices superior to this plane were taken from the superior FDK reconstructed image, while slice inferior to this plane were taken from the inferior FDK reconstructed image. In the case of the maximum 20 cm spacing, we combined the entire image space from each reconstruction.

III. Results

Figure 1 shows both the combined FDK and MLEM reconstructions of the Defrise-style disk phantom from the circular scans spaced 10 cm apart. The design of this phantom makes it particularly susceptible to cone-angle artifacts in reconstructed images. Figure 1a illustrates cone-angle artifacts in the middle and at the axial ends of the image. Comparing this to the MLEM reconstruction in Figure 1b, the uniform disks appear much sharper, especially in the regions where the FDK image suffers from cone-angle artifacts. The bottom row of Figure 1 demonstrates that the slice through the central high-density uniform disk in the phantom appears to have values which are more correct in the MLEM reconstruction than that of the stacked FDK.

With a more extreme 20 cm offset between the two circular scans, the difference between FDK and MLEM is even more pronounced as shown in Figure 2. In the mid-sagittal slices shown in the top row, the FDK reconstruction in Figure 2a suffers from extreme cone-angle artifacts. Though some of the disks in the center of the MLEM reconstruction in Figure 2b are somewhat distorted, it does not exhibit the same degree of degradation. In the mid-transverse slices shown on the bottom row, the MLEM slice in Figure 2d shows some non-uniformity. However, it is much better than the FDK reconstruction shown in Figure 2c.

Using the voxelized version of the disk phantom, we compared the MLEM and FDK reconstructions to the truth with root-mean-square error (RMSE). Figure 3 shows a comparison of the RMSE of MLEM and FDK reconstructions as a function of spacing. This plot indicates that for all of the circular scan spacings simulated in this study, MLEM consistently outperforms the stacked FDK reconstruction. Furthermore, as the spacing increases, the RMSE of FDK also increases. For MLEM, the RMSE increases much more slowly than FDK with increasing spacing.

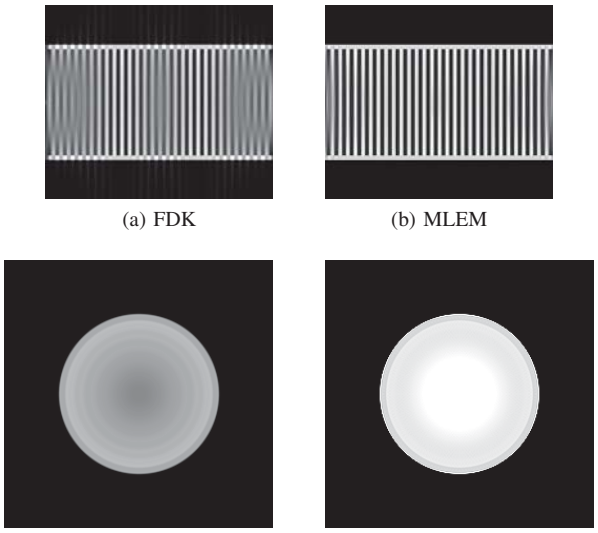
For the XCAT phantom reconstruction, Figure 4 shows slices of both the MLEM and the FDK stacked reconstructed image volumes from the circular scans spaced 10 cm apart. Comparing the two reconstruction methods, the overlap of the image space seems sufficient to avoid corruption of the central plane from cone-angle artifacts. It is worth noting that in the MLEM reconstruction, the edges of structures such as the skull are sharper. Also, the optic chiasm can be distinguished in the mid-transverse MLEM image shown in Figure 4d whereas it cannot be seen in the corresponding stacked FDK reconstruction shown in Figure 4c.

Figure 5 shows reconstructions of the XCAT with the circular scans spaced at the maximum 20 cm apart. Figures 5a and 5c show FDK is unable to accurately reconstruct the central region between the two circular scans. At this maximum spacing, the worst-case manifestation of cone-angle artifacts severely degrade the quality of the image's central volume. MLEM however, does a very good job of handling this central region as shown in Figures 5b and 5d. Though there are some streaks that can be seen in the central portion of the MLEM image, it is a clear improvement to the extensive image degradation seen in the stacked FDK images.

IV. Conclusion

These results show that iterative reconstruction methods can be very effective in extending the axial FOV. MLEM was much more successful in avoiding the cone-angle artifacts that plagued the analytic FDK reconstruction in the region between

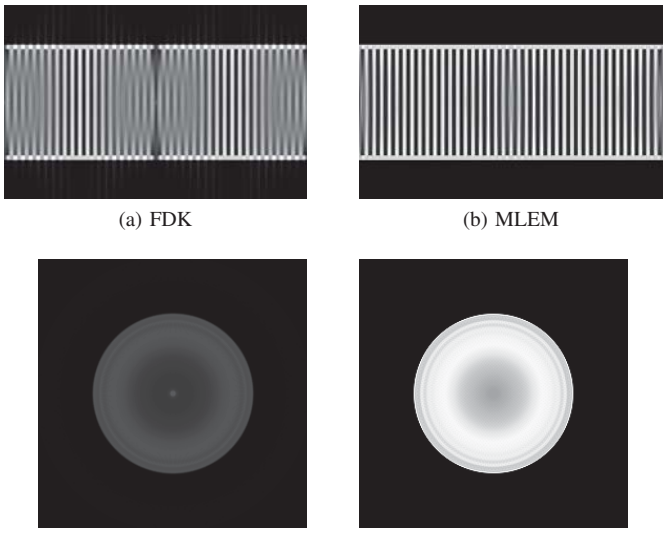
the two circular acquisition planes. It is also clinically relevant that the extended axial FOV obtained with this dual circle acquisition is 40 cm. For the IGRT machine we based our OBI geometry on in this study, the maximum beam coverage is also 40 cm.



(a) FDK (b) MLEM

(c) FDK (d) MLEM

Fig. 1. Defrise-disk phantom reconstructions. The display window is [0, 0.3]. Top: mid-sagittal slice using two independent FDK reconstructions (a) and 200 iterations of direct MLEM reconstruction (b). Bottom: mid-transverse slices from the FDK (c) and MLEM (d) volumes.



(a) FDK (b) MLEM

(c) FDK (d) MLEM

Fig. 2. Defrise-disk phantom reconstructions. The display window is [0, 0.3]. Top: mid-sagittal slice using two independent FDK reconstructions (a) and 200 iterations of direct MLEM reconstruction (b). Bottom: mid-transverse slices from the FDK (c) and MLEM (d) volumes.

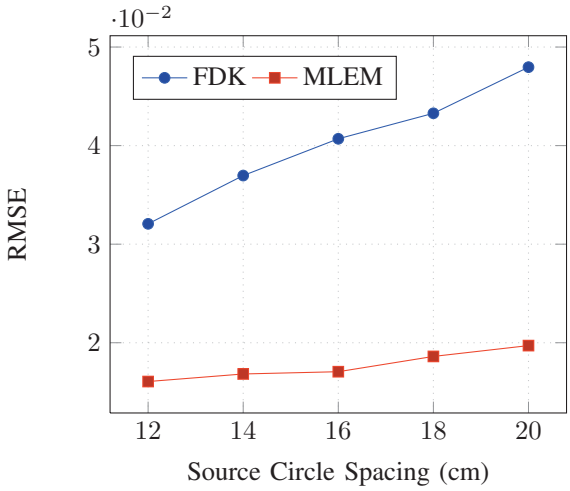
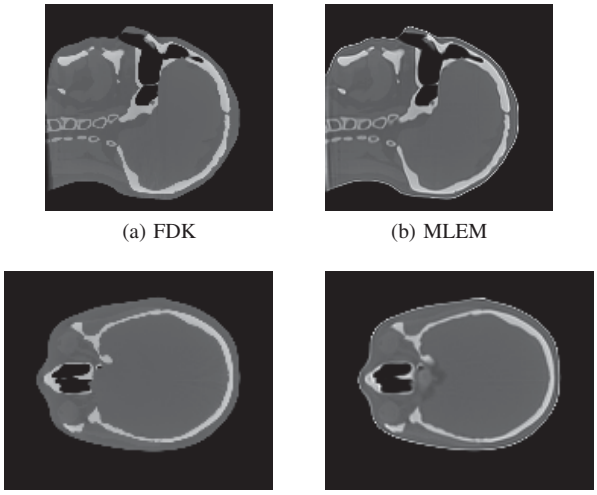


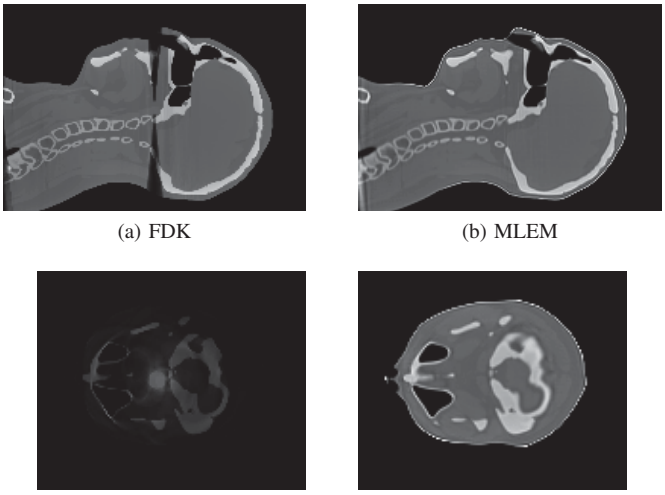
Fig. 3. RMSE comparison of MLEM and FDK disk reconstructions with a voxelized version of the disk phantom



(a) FDK (b) MLEM

(c) FDK (d) MLEM

Fig. 4. XCAT phantom reconstructions from two circular scans 10 cm apart. The display window is [0.1, 0.4]. Top: mid-sagittal slice using two independent FDK reconstructions (a) and 200 iterations of direct MLEM reconstruction (b). Bottom: mid-transverse slices from the FDK (c) and MLEM (d) volumes.



(a) FDK (b) MLEM

(c) FDK (d) MLEM

Fig. 5. XCAT phantom reconstructions from two circular scans 20 cm apart. The display window is [0.1, 0.4]. Top: mid-sagittal slice using two independent FDK reconstructions (a) and 200 iterations of direct MLEM reconstruction (b). Bottom: mid-transverse slices from the FDK (c) and MLEM (d) volumes.

REFERENCES

[1] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Problems*, vol. 25, no. 12, p. 123009, Dec. 2009. [Online]. Available: http://iopscience.iop.org/0266-5611/25/12/123009

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, Jun. 1984. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-1-6-612

[3] E. Pearson, X. Han, X. Pan, and C. Pelizzari, "Iterative reconstruction for axial field of view extension in radiotherapy cone-beam CT," Valencia, Spain, Oct. 2011.

[4] E. Pearson, S. Cho, C. Pelizzari, and X. Pan, "Non-circular cone beam CT trajectories: A preliminary investigation on a clinical scanner," in *2010 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, Nov. 2010, pp. 3172 –3175.

[5] J. Müller-Merbach, "Simulation of x-ray projections for experimental 3D tomography," Image Processing Laboratory, Department of Electrical Engineering, Linkoping University, SE-581 83, Sweden, Tech. Rep. LiTH-ISY-R-1866, Mar. 1996.

[6] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, no. 8, pp. 3800–3808, Aug. 2008, PMID: 18777939 PMCID: PMC2809711. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809711/

Noise Energy Estimation Based on the Sinogram and its Application to the Regularization Parameter Selection for Statistical Iterative Reconstruction

Ti Bai, Xuanqin Mou, Qiong Xu and Yanbo Zhang

***Abstract*—Regularized statistical CT reconstruction can yield a high-quality image due to the incorporation of physical constraints, accurate system model and stochastic property. Nevertheless, regularization parameter that is related to noise energy of measurements needs to be tuned properly so as to acquire a good reconstruction. Unfortunately, little attention has been paid to the estimation of noise . In this paper, we proposed a novel method to estimate noise energy of measurements based on the Fourier properties of sinogram. In addition, we applied the noise energy estimated to adaptive selection of regularization parameter. Numerical results indicated that our method can estimate the noise energy precisely and robustly, by the way, algorithm for adaptive selection of regularization parameter also worked very well by making use of the noise energy estimated by us.**

***Index Terms*—Noise estimation, Fourier properties, sinogram, regularization parameter, ASD-POCS**

I. Introduction

BY formulating the statistical characteristics of projection data, statistical iterative reconstruction (SIR) can obtain better reconstruction quality compared with conventional methods such as filtered back projection (FBP) and algebraic iterative technique (ART). However, the unregularized problem is poorly conditioned or even underdetermined, so regularization term is required to ensure a stable solution. Numerous regularizers have been proposed, which include smooth regularizers [1] and sparsity-promoting ones [2]–[4]. Among them, sparsity regularization based on dictionary learning shows promising results in terms of low dose CT reconstruction [3], [4]. In general, regularization parameter that balances data fidelity term and regularization term needs to be tuned carefully in order to obtain a high-quality reconstruction.

Various regularization parameter selection methods have been proposed including the discrepancy principle (DP) [5], generalized cross validation (GCV) [6], the L-curve method [7] and Stein's unbiased risk estimate (SURE) [8], [9]. The use of DP and SURE require the knowledge of noise level which is not necessary for the other two. Nevertheless, the L-curve method can be computationally intensive and sensitive to curvature estimation. As for GCV which is especially suitable for linear problem, it will become computationally involved when applied to nonlinear problems. By the way, the ASD-POCS algorithm proposed by Sidky E. Y. *et al.* [2] is a special case of superiorization methodology which is proposed by G.T. Heman *et al.* [10], both of them utilize the noise energy of avaiable data. By and large, a good estimation of noise energy in the data will benefit a lot for the choice of regularization parameter as mentioned above. Unfortunately, to the best of our knowledge, little attention has been paid to this subject in the field of CT. The nonidentical distribution of noise in raw projection which makes it troublesome to handle maybe a major cause that leads to this phenomenon. A modeling method to estimate the variance of projection data has been proposed in the literature based on theoretical analysis and experimental verification [11], [12]. It is very complicated for that the modeling method works by polynomial fitting. T.Niu *et al.* [13] proposed an appealing method by taking use of the property of Poisson noise, however, it is difficult in practical applications because it must to complete a consecutive flatfield scans. Based on the work of Samuel R. Mazin *et al.* [14] describing Fourier properties of the full fan-beam sinogram, in this paper, we proposed a novel method to estimate the noise energy in the sinogram by making use of the Fourier properties of the fan-beam sinogram. It is worthy to mention that we just need the sinogram which is not the case of [13], besides, the computation is also very cheap, one Fast Fourier Transform (FFT) is enough.

In the next section, we will derivate the method for noise estimation in details and describe one of its applications for adaptive parameter selection in the dictionary learning based statistical iterative reconstruction (SIR). In the third section, numerical results will be shown to validate our method. Finally, we will discuss some issues in this study and conclude this paper.

II. Methods

A. Noise Estimation

In this paper, we supposed that after the effective data correction algorithms, including scatter and beam hardening corrections, Gaussian noise on the projections is dominant. Assuming the following Gaussian statistical model for measurements of line integral:

$$p = p^0 + n, \tag{1}$$

we denoted by $p^0(\gamma, \beta)$ the noise-free ray indexed by (γ, β), and $p(\gamma, \beta)$ is the corresponding noisy counterpart, n represents the noise that is signal-dependent. In the following, we

This work was supported in part by National Natural Science Foundation of China (NSFC) (No. 61172163), in part by the Research Fund for the Doctoral Program of Higher Education of China (No. 20110201110011).

The authors are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China. (Email: baiti018@xjtu.edu.cn, xqmou@mail.xjtu.edu.cn, xjtuxqiong@gmail.com, yanbozhang007@163.com)

will present the Fourier properties of both the free sinogram and noise. Then, method to estimate the noise energy will be derived by making use of this Fourier properties.

1) Fourier properties of the fan-beam sinogram: For the amplitude of a two dimensional Fourier transform of full fan-beam sinogram that is noise free, one of its noteworthy Fourier properties of the sinogram is that there exists a region with the shape of double-wedge whose energy is approximately zero. For clarity, we present here the results derived in [14].

Considering an object f consists of a delta function locating at distance r_p from the origin and at angle ϕ_p with the x-axis, γ is the cylindrical detector fan angle, β is the view angle, L represents the distance from source to isocenter, then the randon transform of f can be written as follows:

$$p^0(\gamma,\beta) = \delta(\gamma - \gamma_0(\beta)),$$

where

$$\gamma_0(\beta) = \arctan[\frac{\frac{r_p}{L}\cos(\phi_p-\beta)}{1-\frac{r_p}{L}\sin(\phi_p-\beta)}].$$

Taking two dimensional Fourier transform for the randon transform of f, we have

$$\begin{aligned} P(q,k) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\infty}^{\infty}\delta(\gamma-\gamma_0(\beta))e^{-j(q\gamma+k\beta)}\mathrm{d}\gamma\mathrm{d}\beta \\ &\approx e^{-jk(\phi_p+\frac{\pi}{2})}J_k((k-q)\frac{r_p}{L}), \end{aligned} \tag{2}$$

where q,k are the frequency variables corresponding to γ and β respectively, $J_k((k-q)\frac{r_p}{L})$ is a Bessel function of the first kind of order n with the argument $(k-q)\frac{r_p}{L}$. From the property of Bessel function[1], there exists a double wedge region where the energy of the projections is equal to zero, which can be segmented by checking if the frequencies (q,k) satisfy the following condition:

$$\|\frac{k}{k-q}\| > \frac{r_p}{L}. \tag{3}$$

2) Fourier properties of noise in the sinogram: Fourier properties of noise will be derived in a discrete form. Denote $n(\gamma,\beta)$ the zero-mean real Gaussian random variable with variance $\sigma^2(\gamma,\beta)$ that relates to the detector element (γ,β), the discrete Fourier transform of noise can be written as:

$$N(q,k) = \sum_{\gamma}\sum_{\beta} n(\gamma,\beta)e^{-j(q\gamma+k\beta)}.$$

As $N(q,k)$ is a zero-mean *complex* Gaussian random variable, we have

$$\begin{aligned} \mathrm{var}(N(q,k)) &= \sum_{\gamma}\sum_{\beta}\sigma^2(\gamma,\beta) \\ &= \mathrm{E}(N(q,k)\times\overline{N(q,k)}) \\ &= \mathrm{E}(\|N(q,k)\|_2^2), \end{aligned} \tag{4}$$

where $\overline{N(q,k)}$ denotes the conjugate, $\mathrm{var}(N(q,k))$ is the variance of $N(q,k)$. (4) reveals that expectation of square of the module of any Fourier coefficient for noise is equal to the summation of the variance of noise.

[1]The Bessel function drops dramatically and tends to zero when the argument is less than its order.

3) Noise estimation based on the Fourier properties: On the basis of the Law of Large Numbers and the definition of noise energy and variance, we have

$$\begin{aligned} \epsilon^2 &= \sum_{\gamma}\sum_{\beta}(p(\gamma,\beta)-p^0(\gamma,\beta))^2 \qquad (5) \\ &= \mathrm{E}[\sum_{\gamma}\sum_{\beta}(p(\gamma,\beta)-p^0(\gamma,\beta))^2] \\ &= \sum_{\gamma}\sum_{\beta}\sigma^2(\gamma,\beta) \\ &= \mathrm{E}(\|N(q,k)\|_2^2) \qquad (6) \end{aligned}$$

Just as described above, there exists a approximately zero-energy region in the two dimensional Fourier transform of the free fan-beam sinogram while expectation of square of the module in any point for the case of noise is equal and its value is exactly noise energy ϵ^2. As for the linearity of Fourier transform and the Law of Large Number once again, we can estimate the noise energy as follows:

1) take Fourier transform of the noisy sinogram,
2) extract the region where the signal energy is approximately zero while the noise energy is reserved,
3) sum up the square of module in the extracted region and average it, the result is just the noise energy we estimate.

B. Adaptive Regularization Parameter Selection

In order to demonstrate the efficiency to select the regularization parameter when the noise energy is avaiable by our method, we designed a special algorithm that considered the dictionary learning based sparsity regularization as *a prior* information. Following the spirit of [2], [10], we selected the optimal value of regularization parameter that should yield a solution which is more tolerance-compatible than the phantom. Image reconstruction can be formulated as follows:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x},\boldsymbol{\alpha}} \{&\frac{1}{2}(\mathbf{Ax}-\mathbf{y})^{\mathbf{T}}\mathbf{C}^{-1}(\mathbf{Ax}-\mathbf{y}) + \\ &+\beta(\sum_s\|\mathbf{E}_s\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}_s\|_2^2+\lambda\sum_s\|\boldsymbol{\alpha}_s\|_0)\}, \end{aligned} \tag{7}$$

where $\mathbf{x}\in\mathbf{R}^N$ is the attenuation map to be reconstructed, we denoted by $\mathbf{y}\in\mathbf{R}^M$ the M-length data vector. $\mathbf{A}=\{a_{ij}\}, i\in\{1,\cdots,M\}, j\in\{1,\cdots,N\}$ is the system matrix, $\mathbf{C}=\mathrm{diag}\{\sigma_i^2\}$ is the $M\times M$ covariance matrix.[2] The operator $\mathbf{E}_s$ extracts the s^{th} patch from the current reconstruction, and $\mathbf{D}$ is the pretrained dictionary under which the sparse representation of the s^{th} patch is $\boldsymbol{\alpha}_s$. β is the regularization parameter that is to be adaptively selected in this study to control the tradeoff between data fidelity term and regularization term, and λ is the parameter for sparse coding that is usually set to a constant.

Update of the attenuation map $\mathbf{x}$ are due to both data fidelity term and regularization term. In order to reserve medically relevant features, overfitting is desirable [10] for that the parameter β which results in a tolerance-compatible solution will oversmooth the reconstruction. In this paper, we define the feasible set as $S=\{\mathbf{x}\mid 0.9\epsilon^2<\|\mathbf{Ax}-\mathbf{y}\|_2^2<0.92\epsilon^2\}$, in

[2]For simplicity, we used $\sigma_i^2=e^{y_i}$ in this paper.

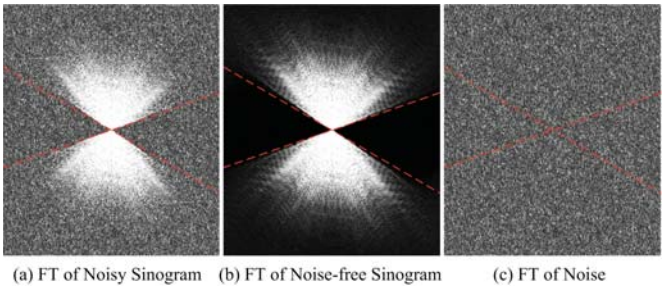


Fig. 1. Square of the module of Fourier cofficients. (a)-(c) corresponds to noisy sinogram, noise-free sinogram and pure noise, respectively. It can be clearly seen that there exists a zero-energy region (inside of the red cross) for that of noise-free sinogram, while for the case of pure noise, it is quite stationary.

other words, if the current image belonged to S, we regarded the current β as a potential optimal one and remain unchanged; otherwise, if the current image satisfied $\|\mathbf{Ax}-\mathbf{y}\|_2^2 > 0.92\epsilon^2$ and the ratio of update due to regularization term to that due to data fidelity is greater than $r_{\max}$, it indicated that the current β was too large, and we should reduce it by a constant fraction β_c; on the contrary, we should increase it by $\frac{1}{\beta_c}$, repeat this process until the algorithm converges. In this paper, we initialized β by 1, and set $r_{\max}$, β_c as 0.98, 0.95 respectively, the target error λ for sparse coding is fixed with 1×10^{-6}.

III. Numerical Experiments

A. Results for Noise Estimation

To validate our algorithm, we presented numerical results of noise estimation for two kinds of 672×580-view sinogram which were numerically generated from two 2-D simulated human phantom with SIEMENS Somatom Sensation geometry. Various Poisson noise of a range of number of incident photons from 1×10^4 to 1×10^6 respectively were superimposed into the raw noise-free projections to synthesize kinds of dose data. Following the spirit of Section.II-A3, we took 2-dimensional Fourier transform for the noisy sinogram, then extracted the zero-energy region and averaged it to estimate the noise energy.

Figure.1(a)-(c) illustrated results of Fourier tranform corresponding to noisy sinogram, noise-free sinogram and pure noise, respectively. The zero-energy region of which we made use to estimate the noise energy can be seen clearly in Fig.1(b). By averaging this zero-energy region, we can estimate the noise energy of some noisy sinogram. Fig.2 shows that the estimations are quite robust for different kinds of noise level regardless of different phantoms. The relative root mean square error(RRMSE) for the abdomen phantom, see Fig.2(a), is 1.8%, while it is 0.64% for the human thorax phantom, see Fig.2(b).

B. Results for Adaptive Regularization Parameter Selection

We used the 512×512 2-D abdomen phantom, illustrated in Fig.2(a), to numerically generate two 672×580-view sinograms with SIEMENS Somatom Sensation geometry corresponding to a monoenergetic source with 5×10^4 and 5×10^5 incident photons per ray, respectively. Then, we estimated the noise energy of these two sinograms by our method. The estimated values were 75.5886 and 7.5162 corresponding to the case of 5×10^4 and 5×10^5 incident photons per ray while their real value were 75.0190 and 7.4999, respectively.

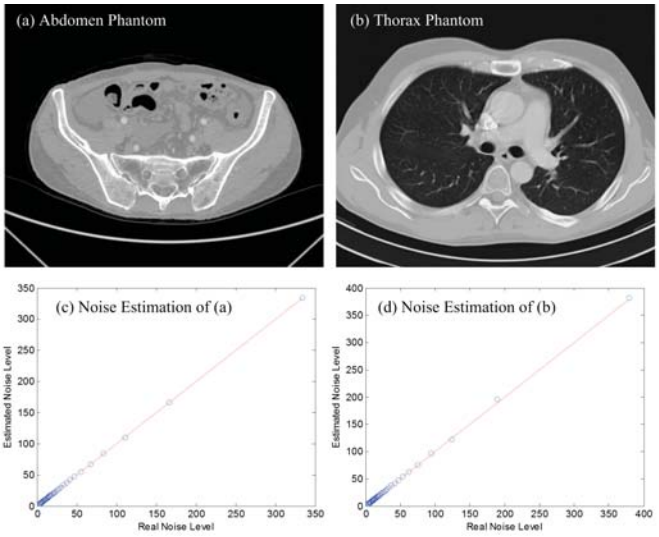


Fig. 2. Results of noise estimation for different noise level and different phantoms. (a) and (b) are noise-free phantom corresponding to abdomen and human thorax, (c) and (d) are results of noise estimation for phantom of abdomen and human thorax, respectively. The blue circle represents the value of real noise energy and its estimated counterpart. Analytic form of red line is $y=x$. It indicated that the method for noise estimation was quite stable. For (c), the RRMSE was 1.8%, while it was 0.64% for (d).

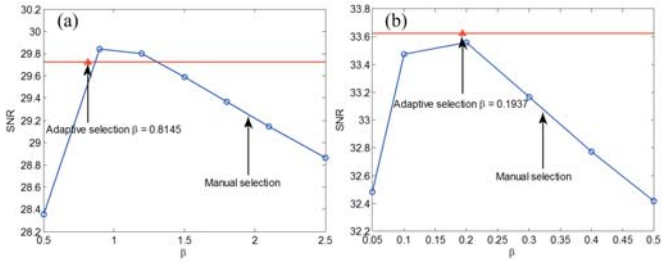


Fig. 3. Plot of SNR as a function of regularization parameter β selected manually by trial and error. (a) and (b) are the case of 5×10^4 and 5×10^5 incident photons, respectively. The plots indicated that β selected adaptively can always arrive at the approximately optimal value for different noise level in the sense of SNR.

With the estimated noise energy, we selected the regularization parameter adaptively according to Section.II-B. Fig.3(a)-(b) plotted results of adaptive regularization parameter selection corresponding to the case of 5×10^4 and 5×10^5 incident photons, respectively, and indicated that we always can select the approximately optimal β adaptively in terms of SNR. To support our arguments further, we displayed reconstructions for these two cases as shown in Fig.4. The images generated using different β were illustrated in the 2^{th}-4^{th} column. The first column were reconstructions of adaptive algorithm while the third column were the highest-quality reconstructions by trial and error. It can be seen that images reconstructed by adaptive algorithm were very similar to those reconstructed by trial and error in the sense of visualization. To obtain images similar to those reconstructed adaptively, β needs to be tuned carefully in a large range when the noise energy changed. For instance, β needed to increase by a factor of 4.5 when the dose reduced 10 times.

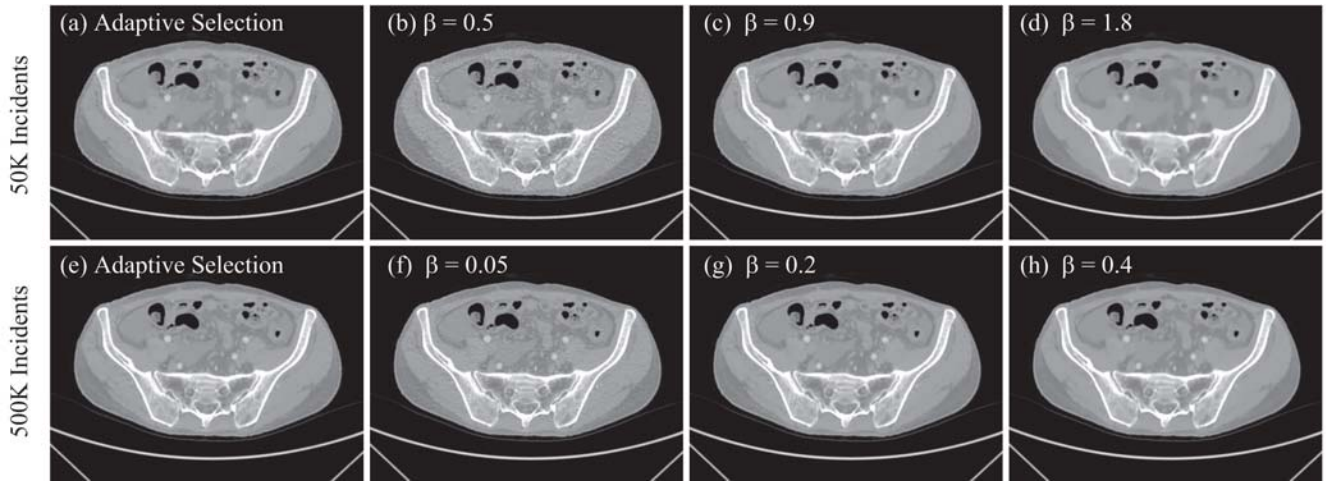


Fig. 4. CT images reconstructed with different β for different noise level. The top row and the bottom row corresponds to the case of 5×10^4 and 5×10^5 incident photons. Images of the first column were reconstructed adaptively while the others were reconstructed using different β, among which, the third column had the highest quality in the sense of visualization. It can be seen that images reconstructed by adaptive algorithm were very similar to the highest quality reconstruction by trial and error in the sense of visualization.

IV. Discussion and Conclusion

As demonstrated above, the proposed method can estimate noise energy accurately and stably. Besides, we also applied the estimated value to the adaptive selection of regularization parameter and obtained a reconstruction similar to the highest-quality image reconstructed by trial and error in the sense of visualization. Nevertheless, there still exist some problems needed to be studied more deeply. Firstly, the filter used to extract the zero-energy region needs to be designed carefully due to the gradual transitions out of the zero-energy region in the transformed sinogram. In practice, we always extract the signal in the region which is smaller than that defined by (3). Secondly, although we demonstrated that the proposed method can estimate the noise energy accurately and stably when the poisson noise in the raw data is dominant[3] in this study, it is still worthy to make it clear that whether the proposed algorithm works well as usual when the data is not corrected. Last but not least, the algorithm to select the regularization parameter needs a deeper discussion. As mentioned above, the reconstructions happening to be tolerance-compatible will be oversmoothed, which will hamper the observation of fine image details. Algorithms with better parameter selection framework should be explored.

In summary, we proposed a novel method to estimate the noise energy and numerical results indicated that this method is quite accurate and stable. Moreover, with the help of the estimation of noise energy, we selected the regularization parameter adaptively and produced promising results suggesting that images reconstructed by adaptive algorithm are parallel to those reconstructed by trial and error.

References

[1] I. Elbakri and J. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 2, pp. 89–99, 2002.

[2] E. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in medicine and biology*, vol. 53, no. 17, p. 4777, 2008.

[3] Q. Xu, H. Yu, X. Mou, and G. Wang, "Dictionary learning based low-dose x-ray ct reconstruction," in *11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2011, pp. 259–63.

[4] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray ct reconstruction via dictionary learning," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 9, pp. 1682–1697, 2012.

[5] H. Engl, "Discrepancy principles for tikhonov regularization of ill-posed problems leading to optimal convergence rates," *Journal of optimization theory and applications*, vol. 52, no. 2, pp. 209–215, 1987.

[6] G. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[7] P. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992.

[8] Y. Eldar, "Generalized sure for exponential families: Applications to regularization," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 471–481, 2009.

[9] R. Giryes, M. Elad, and Y. Eldar, "The projected gsure for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–422, 2011.

[10] G. T. Herman, E. G. no, R. Davidi, and Y. Censor, "Superiorization: An optimization heuristic for medical physics," *Medical Physics*, vol. 39, no. 9, pp. 5532–5546, 2012. [Online]. Available: http://link.aip.org/link/?MPH/39/5532/1

[11] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, and J. Manzione, "An experimental study on the noise properties of x-ray ct sinogram data in radon space," *Physics in medicine and biology*, vol. 53, no. 12, p. 3327, 2008.

[12] S. Tang and X. Tang, "Statistical ct noise reduction with multiscale decomposition and penalized weighted least squares in the projection domain." *Medical physics*, vol. 39, no. 9, p. 5498, 2012.

[13] T. Niu and L. Zhu, "Accelerated barrier optimization compressed sensing (abocs) reconstruction for cone-beam ct: Phantom studies." *Medical physics*, vol. 39, no. 7, p. 4588, 2012.

[14] S. R. Mazin and N. J. Pelc, "Fourier properties of the fan-beam sinogram," *Medical Physics*, vol. 37, no. 4, pp. 1674–1680, 2010. [Online]. Available: http://link.aip.org/link/?MPH/37/1674/1

[3] Virtually such is the case after the effective data correction algorithms, including scatter and beam hardening corrections

Fast Variance Prediction for Iterative Reconstruction of 3D Helical CT Images

Stephen M. Schmitt and Jeffrey A. Fessler

***Abstract*—Fast variance prediction for iteratively-reconstructed helical CT images is useful for analysis of resulting images and potentially for dynamic dose adjustment during a scan. Previous methods require impractical computation times to approximate the image variance; other methods are able to approximate variance quickly but only for specific CT geometries, excluding 3D helical CT. In this paper we present an extension of these previous fast methods to predict the variance of iteratively reconstructed images for third-generation 3D helical CT scans. We compare this method in computation time and error to the empirical variance derived from multiple simulated reconstruction realizations.**

I. Introduction

Iterative reconstruction (IR) methods for computed tomography are receiving increased attention for their improved resolution and noise properties compared to FBP [7]. However, the statistical properties of IR reconstructions are difficult to compute compared to FBP. Closed-form but computationally intractable expressions exist [1] for the mean and covariance matrix of the reconstruction when the weighting matrix $\mathbf{W}$ and covariance of the projections are given, so faster prediction methods are desirable.

Prior work has exploited approximate local shift-invariance to develop FFT-based approximations for the variance map of the image, i.e., the diagonal of the covariance matrix, for arbitrary system geometries [5]. Unlike empirical methods, which can only be used to find the variance map of the entire image simultaneously, these FFT-based methods can approximate the variance of one specific voxel of interest at a time. However, these FFT-based methods are computationally intensive; they are useful for theoretical analysis but require projection and back-projection of each voxel of interest and are unsuitable for producing a variance map for a whole volume. There are methods for 2D fan-beam [9], 3D step-and-shoot [10], and 3D axial CT [6] that make further approximations to greatly reduce the computational load of this method and make it suitable for predicting variance maps for an entire volume. None of these methods, though, apply directly to 3D helical CT.

In this paper, we adapt [6] to the problem of predicting approximate variance maps for iterative reconstruction of 3D helical CT scans. Like this prior work, the computational cost of the variance approximation is reduced by several orders of magnitude compared to empirical estimation or the FFT-based method in [5] for CT.

This work was supported in part by NIH grant R01 HL-098686 and by equipment donations from Intel.

Stephen M. Schmitt and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (smschm@umich.edu, fessler@umich.edu)

II. Methods

A reconstruction method using a weighted least squares data-fit term using log-sinogram observations $\mathbf{y}$, system matrix $\mathbf{A}$ and a regularization term R is given by

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \frac{1}{2}||\mathbf{y}-\mathbf{A}\mathbf{x}||^2_{\mathbf{W}} + \frac{\alpha}{2}R(\mathbf{x}). \tag{1}$$

With a weighting matrix $\mathbf{W} = \operatorname{cov}(\mathbf{y})^{-1}$ and assuming that the minimization algorithm is iterated until convergence, the covariance matrix of $\hat{\mathbf{x}}$ in (1) is approximately [1]:

$$\operatorname{cov}(\hat{\mathbf{x}}) \approx (\mathbf{H}+\alpha\nabla^2 R(\hat{\mathbf{x}}))^{-1}\mathbf{H}(\mathbf{H}+\alpha\nabla^2 R(\hat{\mathbf{x}}))^{-1}. \tag{2}$$

If $R(\mathbf{x}) = \sum_i \psi([\mathbf{C}\mathbf{x}]_i)$ for a matrix $\mathbf{C}$ and a convex penalty function ψ that is twice-differentiable in an open set containing 0 with $\psi(0) = \psi'(0) = 0$, $\psi''(0) = 1$, then

$$\nabla^2 R(\mathbf{x}) = \mathbf{C}^{\mathrm{T}}\ddot{\Psi}(\mathbf{x})\mathbf{C}, \tag{3}$$

where $\ddot{\Psi}(\mathbf{x})$ is a diagonal matrix with $[\ddot{\Psi}(\mathbf{x})]_{jj} = \psi''([\mathbf{C}\mathbf{x}]_j)$. With a sufficiently large α, we would expect that, for most voxels, $[\mathbf{C}\mathbf{x}]_j$ is small and in the twice-differentiable region of ψ and therefore, that $\ddot{\Psi}(\hat{\mathbf{x}}) \approx \mathbf{I}$ is a valid approximation except near edges between regions of different attenuation coefficients in the image. Making this substitution transforms (2) into

$$\operatorname{cov}(\hat{\mathbf{x}}) \approx (\mathbf{H}+\alpha\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{H}(\mathbf{H}+\alpha\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}, \tag{4}$$

where $\mathbf{H} \triangleq \mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$. However, direct computation of this matrix is not computationally tractable.

A. Prior work

In [6], we define a continuous-frequency response operator *local* to the jth voxel:

$$\left(\mathscr{F}^3_{j,\mathrm{cont}}\{\mathbf{x}\}\right)(\vec{\nu}) = \sum_{\ell=1}^{|\vec{N}|} \mathbf{x}_\ell \exp\left(-\imath 2\pi\vec{\nu}\cdot(\vec{n}_\ell - \vec{n}_j)\right), \tag{5}$$

where $\vec{n}_j$ is the position, in 3 integer coordinates, of the voxel j. We show that the variance of this voxel j can be estimated by:

$$\operatorname{var}(\hat{\mathbf{x}}_j) \approx \int_{[-\frac{1}{2},\frac{1}{2}]^3} \frac{H_j(\vec{\nu})}{(H_j(\vec{\nu})+\alpha R(\vec{\nu}))^2}\,\mathrm{d}\vec{\nu}, \tag{6}$$

where $H_j \triangleq \mathscr{F}^3_{j,\text{cont}}\{\mathbf{H}_{*j}\}$, representing the frequency response of projection, weighting, and back-projection, and $R \triangleq \mathscr{F}^3_{j,\text{cont}}\{[\mathbf{C}^\mathrm{T}\mathbf{C}]_{*j}\}$, representing the frequency response of the regularizer when $[\mathbf{Cx}]_j \approx 0$.

We also show that $H_j(\vec{\nu})$ can be written as $H_j(\vec{\nu}) \approx K \cdot J(\vec{\nu}) \cdot E_j(\Phi)$. Here, J depends only on the spatial frequency and not the image or voxel location. E_j is dependent on the image and voxel location but only depends on the spatial frequency via its angle Φ in cylindrical coordinates (ρ, Φ, ν_3). When H_j is specified in this form, (6) can be rewritten in a single-integral form:

$$\operatorname{var}(\hat{\mathbf{x}}_j) \approx \alpha^{-1} \int_0^{2\pi} F(\Phi, \alpha^{-1} K E_j(\Phi))\, \mathrm{d}\Phi, \tag{7}$$

where

$$F(\Phi, \gamma) \triangleq \int_0^{\rho_{\max}(\Phi)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\gamma \cdot J(\rho, \Phi, \nu_3) \cdot \rho}{(\gamma \cdot J(\rho, \Phi, \nu_3) + R(\vec{\nu}))^2}\, \mathrm{d}\nu_3\, \mathrm{d}\rho. \tag{8}$$

There is no closed form for (8), but we can numerically integrate and tabulate it for many values of Φ and γ, independently of the image or weighting matrix, for a given CT geometry and regularizer. In doing so, variance estimation via (7) is simply a one-dimensional numerical integration of values looked up in a pre-computed table of F.

We can import much of the derivation in [6] to apply to helical CT instead of axial CT. In particular, we can still approximate $H_j(\vec{\nu}) \approx K \cdot J(\vec{\nu}) \cdot E_j(\Phi)$, where for helical CT one can show:

$$K = \Delta_x^3 \Delta_z D_{sd}^2 / D_{s0}^2 \Delta_s \Delta_t \Delta_\beta \tag{9}$$

$$J(\rho, \Phi, \nu_3) = \frac{\operatorname{sinc}(\rho\cos\Phi)^2 \operatorname{sinc}(\rho\sin\Phi)^2 \operatorname{sinc}(\nu_3)^2}{\rho} \tag{10}$$

$$E_j(\Phi) = \sum_{\beta \in \mathscr{B}_j(\Phi)} \frac{D_{s0}^2 \mathring{w}_{\beta,j}}{D_{\beta,j}\sqrt{D_{s0} - r_j^2 \cos^2(\phi_j - \Phi)}}. \tag{11}$$

Here, Δ_x is the spacing between voxels in the x and y directions; Δ_z is the spacing between voxels in the z direction; Δ_s, Δ_t are the spacings between pixels on the detector in the s and t directions; Δ_β is the spacing (in radians) of detector angles between views; D_{sd} is the distance from the x-ray source to the detector; D_{s0} is the distance from the source to the isocenter; $D_{\beta,j}$ is the distance from the source to the voxel j when the source is at angle β; r_j is the distance from the isocenter to voxel j; ϕ_j is the angle of voxel j when represented in cylindrical coordinates. All distances given above ignore the z-coordinate; all points are projected into the xy-plane before calculating distances. The only term dependent on the object is $\mathring{w}_{\beta,j}$, which is discussed further in the next section.

B. Modification for helical CT

The items changed by the transition to helical CT are $\mathscr{B}_j(\Phi)$, which is the set of source angles β that solve

$$r_j \cos(\phi_j - \Phi) = D_{s0} \cos(\beta - \Phi), \tag{12}$$

and $\mathring{w}_{\beta,j}$. The term $\mathring{w}_{\beta,j}$ is the element of the statistical weighting matrix $\mathbf{W}$ corresponding to the location on the detector where a ray from the source at angle β passing through the voxel j lands (or 0, if this ray does not land on the detector).

Equation (12) is not changed by the transition to helical CT, but the values of β that solve it are different. The solutions are the set of source angles for which the ray passing through voxel j is perpendicular to the frequency vector $\vec{\nu}$, where the ray and frequency vector are *both* projected into the xy-plane.

For axial CT, the set $\mathscr{B}_j$ is given by:

$$\mathscr{B}_j(\Phi) = \{\beta^+, \beta^-\} = \left\{\Phi \pm \arccos\left(\frac{r_j}{D_{s0}} \cos(\phi_j - \Phi)\right)\right\}. \tag{13}$$

This covers all of the solutions in one turn, which covers a maximum range of 2π. For helical CT with an arbitrary starting and ending angles $\beta_{\min}$, $\beta_{\max}$,

$$\mathscr{B}_j(\Phi) = \left\{\Phi \pm \arccos\left(\frac{r_j}{D_{s0}} \cos(\phi_j - \Phi)\right) + k2\pi\right\} \cap [\beta_{\min}, \beta_{\max}] \tag{14}$$

for $k \in \mathbb{Z}$. Axial CT is then a special case of (14). Since a large part of the computational cost of our method is finding (11), the change to helical CT increases the cost of our algorithm linearly in the number of turns.

The other quantity, $\mathring{w}_{\beta,j}$, is unchanged except that the lookup procedure is computed for helical CT instead of axial CT.

III. Results

To evaluate our prediction for the variance map, we compared it to the variance map derived empirically by simulating 93 reconstructions of a $512 \times 512 \times 500$ XCAT phantom (Fig. 1 displays axial, sagittal, and coronal slices) with voxel size $\Delta_x, \Delta_y = 0.977\text{mm}, \Delta_z = 0.625\text{mm}$. The system geometry, based on a third-generation GE helical CT scanner, had $\Delta_s \times \Delta_t = 1.0239 \times 1.0964\text{mm}$ detector element size, $D_{sd} = 949.075\text{mm}$ source-to-detector distance, and $D_{s0} = 408.075\text{mm}$ source-to-isocenter distance. In our simulations, the X-ray source went through 3 rotations of 984 views each, with a pitch of 1. Each reconstruction used an ordered-subset method with 41 subsets for 100 iterations.

The regularization used a first-order differencing matrix $\mathbf{C}$ that considered the 6 face-neighbors of each voxel. These differences were penalized by a Huber cost function:

$$\psi(x) = \begin{cases} x^2/2, & |x| \le \delta \\ \delta|x| - \delta^2/2, & |x| > \delta \end{cases},$$

which satisfies our criteria for cost functions. The value of δ was 200HU. The regularization parameter α was equal to 128. The weighting $\mathbf{W}$ was normalized so that unattenuated rays had a weight of 1. The simulated X-ray beam intensity was 10^5 photons per view. For simplicity, we used a standard edge-preserving regularizer, rather than the modified regularizer considered in [2].

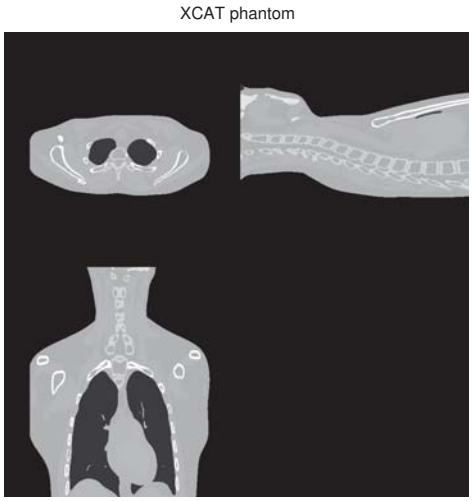


Fig. 1. XCAT phantom (top left is transaxial slice through center of volume; bottom left is center coronal slice; top right is center sagittal slice.)

Figure 2 shows axial, sagittal, and coronal slices of the image of the empirical standard deviation from our simulated reconstructions. Since the results were noisy and the ground truth standard deviation is slowly varying, we blurred the empirical image with a gaussian kernel with a FWHM of 4 voxels each in the x and y directions. Figure 3 shows the corresponding image from our approximation. Since standard deviation varies slowly, we only compute it once per $4\times4\times4$ block and use nearest-neighbor interpolation to fill in the rest. More sophisticated interpolation could be used, but the interpolation error is minimal compared to the intrinsic error of our method. Figure 4 shows the magnitude of the error of our approximated standard deviation. Figure 5 shows both the empirical and approximated standard deviation along a one-dimensional profile through a z-axis of the image. The spike in the empirical map near the end of the axial FOV is due to a suboptimal OS algorithm implementation that is somewhat unstable in regions where the helical sampling is poor. The OS algorithm in [4] would reduce this instability and reduce the empirical variance in the end slices.

The computation time of our method for the entire volume using $4\times4\times4$ downsampling was 1040 CPU-seconds using one core of an Intel Core i7-860 with 16 GB of memory. The empirical reconstructions took a total of 300.8 CPU-days each using one core of an Intel X5650 processor also with 16GB of memory; the range for the individual reconstructions was 2.58 to 3.89 CPU-days.

The axial modulations seen in the coronal and sagittal noise maps were a new phenomena in helical CT variance

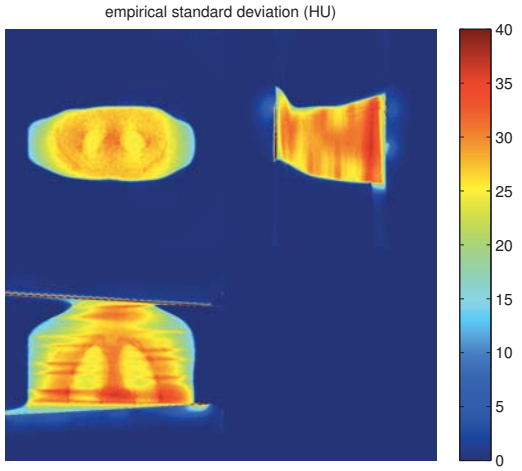


Fig. 2. Empirical standard deviation

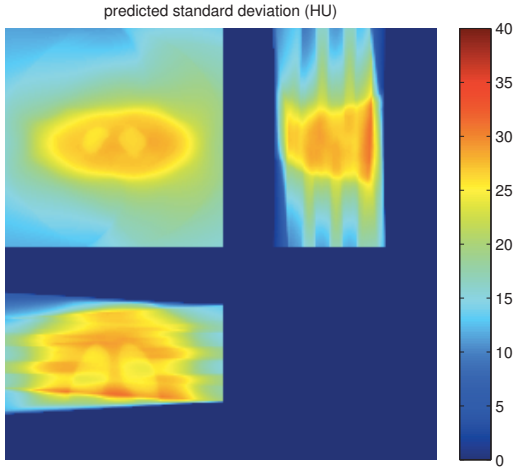


Fig. 3. Predicted standard deviation

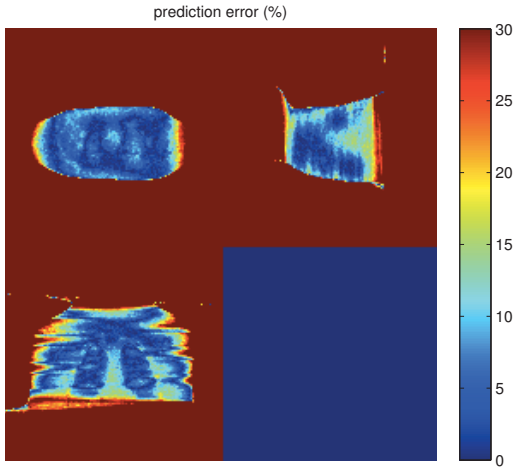


Fig. 4. Error in predicted standard deviation

maps that we had not observed in our previous 3D axial

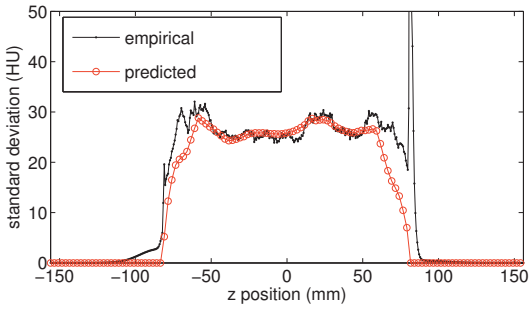


Fig. 5. Empirical & Predicted z-slice through object

CT noise predictions [6]. To help explain this behaviour, we computed a 3D map that shows for each voxel how many rays intersect that voxel. Intuitively, voxels with more intersecting rays are better sampled and thus may have lower variance. Figure 6 shows slices through this ray counting map, and indeed we observe that the sampling pattern influences the predicted and empirical noise maps.

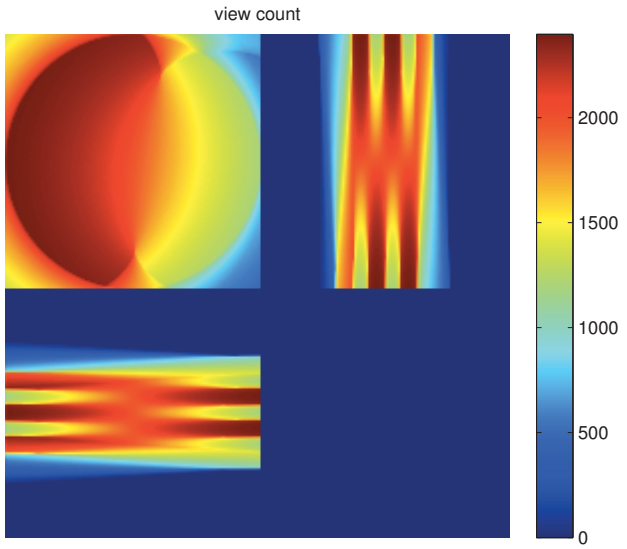


Fig. 6. Count of views that contribute to each voxel's variance prediction

IV. Discussion

The presented methods are able to predict the standard deviation of most voxels in the reconstructed image within an error of 20% in less time than the amount of time empirical measurement takes by a factor of over 10000. The more general (and accurate) approximation using a forward- and back-projection takes 2400 CPU-seconds per voxel (using the same Intel Core i7 above), a factor of over $4 \cdot 10^6$ times as long as our method for one voxel. Whether the tradeoff for time at the expense of accuracy provided by our method is acceptable depends on the application. We also note that our methods would be applicable to axial CT, including short scans, as a special case.

Outside the support of the object there is significant approximation error because our method ignores the non-negativity constraint of the reconstruction. The empirical variance outside the object approaches zero, and so the relative error of our method (which does not go to zero) becomes infinite. An extension to our method could use a pilot reconstruction or masking method (e.g. [3]) to identify external air regions and simply estimate the variance as zero, or use a separate approximation that is more suitable for these regions.

V. Conclusions

In this paper, we have presented a method that is able to approximate the variance of each voxel of a 3D helical CT image reconstructed using a penalized weighted least-squares formulation. This method has a computational cost that is smaller by several orders of magnitude compared to existing variance prediction methods for helical CT, while maintaining a reasonable error within regions of interest.

One direction of future work will be investigating the effect of mismatch between the weighting matrix used for reconstruction and the "optimal" weighting matrix, the inverse of the sinogram covariance. Since the covariance matrix of the sinogram is unknown, in practice we can only approximate it. Knowing the effect of mismatch would also be useful for cases where mismatch is intentional, e.g. [8], to mask out observations known to cause artifacts.

References

[1] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.

[2] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

[3] J. Gregor. Data-driven problem reduction for image reconstruction from projections using gift wrapping. *Nuclear Science, IEEE Transactions on*, 58(3):724–729, 2011.

[4] D. Kim, D. Pal, J-B. Thibault, and J. A. Fessler. Improved ordered subsets algorithm for 3D X-ray CT image reconstruction. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 378–81, 2012.

[5] J. Qi and R. M. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Trans. Med. Imag.*, 18(4):293–305, April 1999.

[6] S. Schmitt and J. A. Fessler. Fast variance computation for quadratically penalized iterative reconstruction of 3D axial CT images. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3287–92, 2012.

[7] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.

[8] K. Zeng, B. De Man, and J-B. Thibault. Correction of iterative reconstruction artifacts in helical cone-beam CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 242–5, 2009.

[9] Y. Zhang-O'Connor and J. A. Fessler. Fast predictions of variance images for fan-beam transmission tomography with quadratic regularization. *IEEE Trans. Med. Imag.*, 26(3):335–46, March 2007.

[10] Y. Zhang-O'Connor and J. A. Fessler. Fast variance predictions for 3D cone-beam CT with quadratic regularization. In *Proc. SPIE 6510, Medical Imaging 2007: Phys. Med. Im.*, pages 65105W:1–10, 2007.

Unified Interpretations of Variants of Simultaneous Algebraic Reconstruction Technique (SART)

Daxin Shi

***Abstract* — In this work, we propose unified interpretations of two variants of the SART algorithms which accommodate weighting schemes. One of the variants is weighted SART and the other one is the weighted least square (WLS) approach. Following the interpretations of the conventional SART, we interpret both variants within the weighted least square solution framework and within the framework of preconditioned gradient descent minimization algorithm. As a consequence of our interpretation, we proposed an alternative derivation of the WLS method without surrogate seeking.**

***Index Terms*—CT reconstruction, weighted SART, iterative reconstruction algorithm**

I. Introduction

Algebraic reconstruction technique (ART) played an important role in the image reconstruction techniques when the first generation of commercial CT scanner entered the market [1-2]. Its history can be even dated back to Kaczmarz [3]. ART possesses a solid foundation of mathematical interpretation, namely, it is a procedure of projection on convex sets to solve a set of large scale linear system equations [4-5]. Unlike the conventional expectation maximization (EM) algorithm [6] which is popular in emission tomographic imaging regime, ART does not necessarily require the positivity constraint on the object function, which provides more flexibility for ART to be applied in other imaging regimes [7]. It has been demonstrated that ART together with particular regularization term, such as total variation, can yield very accurate results in some incomplete data imaging scenario [5]. A disadvantage of the ART is that it is hard to be parallelized due to its ray by ray updating scheme although some effort had been devoted to parallelizing it [8]. The desire of being easily paralleled is due to the fact that in general iterative reconstruction (IR) algorithms such as ART demands much more computational power than the analytic reconstruction methods such as the filtered backprojection (FBP) algorithm. Parallel computation is one way to speed up IR to make it commercially practical.

Simultaneous algebraic reconstruction technique (SART) is an interesting implementation of ART [9]. SART takes many interesting features of the ART, such as no requirement of positivity constraint of the object function. The most attractive feature of SART is that it can be very easily parallelized. Due to the evolution of hardware development, such as the availability of graphics processing units (GPU), the easy parallelizability of SART had become a more important feature. However, the mathematical interpretation of SART was not provided in its original work [9]. Wang and Jiang provided a heuristic derivation of SART [10] and a preconditioned gradient descent interpretation of SART in [11]. The convergence properties of SART had been studied in their subsequent papers [11-12]. Gregor and Benson revisited SART and provided an interesting mathematical interpretation of the algorithm under the framework of a solution to the weighted least square minimization problem [13].

It is generally believed that statistical information should be built into IR which accounts for the real physical imaging condition to yield images of better quality [14]. In their original format, neither ART nor SART took the statistical information into account. Weighted version of these algorithms can be viewed as variants of the original ART and SART. To accommodate the statistical information, Kohler [15] *et. al.* modified the standard ART reconstruction formula with application to PET imaging, where a weight function estimated by means of Gaussian error propagation was introduced. To build the statistical information into the conventional SART, the weighted version of SART had been proposed in Ref. [16]. However, the mathematical interpretation of weighted SART was not provided. The motivation of the weighted SART is that when weights are constant, the algorithm should reduce to its original form of SART [16]. In this work, we will follow the ideas of Gregor and Benson [13] to provide a mathematical interpretation of the weighted SART. We will also provide another interpretation of weighted SART under the framework of gradient descent minimization strategy. Another interesting iterative reconstruction scheme which accounted for statistical information was proposed by Elbakri and Fessler [17]. The interesting part in their algorithm is that they started out from Poisson statistical model in x-ray CT and ended up with a weighted least square minimization problem. With some surrogate seeking, the final reconstruction formula closely resembles the conventional SART algorithm with weighting scheme. Due to this similarity, we are going to propose an alternative derivation without using surrogate and an alternative

D. Shi is with Toshiba Medical Research Institute USA, Inc., Vernon Hills, IL, 60061, USA. E-mail: dshi@ tmriusa.com.

interpretation of this approach.

II. Review: Interpretations of SART

A. Interpretations of SART Algorithm

In this section, we review the mathematical interpretation of the conventional SART proposed in [13]. Consider the following unconstrained minimization problem,

$$x = \arg\min_x \; f(x) = \frac{1}{2}(Ax - p)^T W_{FPJ}(Ax - p), \quad (1)$$

where we have modelled the x-ray CT projection as a linear system of equations

$$Ax = p. \quad (2)$$

The matrix A is referred to as the system matrix, the vector x is the image volume to be reconstructed and the vector p represents the measured projection data. Let $a_{i,j}$ denote the (i, j) entry of the matrix A, x_j be the j^{th} element of the image volume and p_i be the i^{th} datum read from a detector bin. Once the forward projection model is selected, elements of the matrix A are also defined. One of the simplest ways to determine the matrix A is to employ Siddon's forward projection model [18]. The matrix W_{FPJ} is a diagonal matrix which serves as a weighting scheme. The diagonal elements of W_{FPJ} is defined as

$$(W_{FPJ})_{i,i} = \frac{1}{\sum_j a_{i,j}}. \quad (3)$$

One can see the weighting matrix is related to the forward projection model and hence we call it forward projection model weighting (FPJ-weighting) matrix. The necessary optimality condition to Eqn. (1) is that the gradient of the objective function must vanish, resulting to solve the following equation,

$$A^T W_{FPJ} Ax = A^T W_{FPJ} p. \quad (4)$$

To solve Eqn. (4), a preconditioning diagonal matrix C is multiplied on both sides of Eqn. (4)

$$CA^T W_{FPJ} Ax = CA^T W_{FPJ} p, \quad (5)$$

where the diagonal elements of C are defined as

$$(C)_{j,j} = \frac{1}{\sum_i a_{i,j}}. \quad (6)$$

Eqn. (5) suggests an iterative reconstruction algorithm

$$x^{(k+1)} = (I - CA^T W_{FPJ} A)x^{(k)} + CA^T W_{FPJ} p, \quad (7)$$

where the matrix I is the identity matrix. We recognize that Eqn. (7) is exactly SART which takes the following component-wise form

$$x_j^{(k+1)} = x_j^{(k)} + \frac{\sum_i a_{i,j}\left(\frac{p_i - \sum_j a_{i,j} x_j^{(k)}}{\sum_j a_{i,j}}\right)}{\sum_i a_{i,j}}. \quad (8)$$

Another interesting interpretation of SART was proposed in [11-12] within the framework of gradient descent algorithm. The matrix C preconditioned gradient of the cost function in Eqn. (1) is

$$C\nabla f(x) = CA^T W_{FPJ} Ax - CA^T W_{FPJ} p. \quad (9)$$

The usual gradient descent algorithm takes the following form

$$x^{(k+1)} = x^{(k)} - \lambda^{(k)}(CA^T W_{FPJ} Ax^{(k)} - CA^T W_{FPJ} p), \quad (10)$$

where the quantity $\lambda^{(k)}$ is the step size of the k^{th} iteration.

B. A variant of the conventional SART

One can have a variant of the conventional SART by redefining the weighting matrix W_{FPJ} and the preconditioning matrix C. Let the weighting matrix W_{FPJ} be the identity matrix I and redefine the diagonal matrix C as follows

$$(C)_{j,j} = \frac{1}{\sum_i (a_{i,j} \sum_k a_{i,k})}. \quad (11)$$

By these definitions, the component-wise formula of Eqn. (7) is

$$x_j^{(k+1)} = x_j^{(k)} + \frac{\sum_i a_{i,j}\left(p_i - \sum_j a_{i,j} x_j^{(k)}\right)}{\sum_i (a_{i,j} \sum_k a_{i,k})}. \quad (12)$$

A similar preconditioned gradient descent interpretation of Eqn. (12) can be readily proposed using the same arguments as above.

It turns out the discussion of convergence property by using the interpretation in Eqn. (7) is simpler [13]. The sufficient condition for convergence is that the spectral radius of the matrix $I - CA^T W_{FPJ} A$ should be less than 1. The same argument applies to the variant in Eqn. (12). It should be noted that Eqn. (12) was proposed in [17] when the regularization term and the counts weighting scheme are neglected.

III. Interpretation of Weighted Version of SART

To accommodate statistical information, the weighted SART algorithm had been proposed as follows [16]

$$x_j^{(k+1)} = x_j^{(k)} + \lambda^{(k)} \frac{\sum_i a_{i,j} w_i \left(\frac{p_i - \sum_j a_{i,j} x_j^{(k)}}{\sum_j a_{i,j}}\right)}{\sum_i a_{i,j} w_i}, \quad (13)$$

where w_i is the weight for each measured projection datum. In this sense, the weighted SART is also a variant of SART which accounts for weighting scheme. If we follow the interpretation of the conventional SART, the weighted SART can be very easily interpreted as minimizing the following objective function

$$x = \arg\min_x \; f(x) = \frac{1}{2}(Ax - p)^T W(Ax - p) \quad (14)$$

where the diagonal weighting matrix W is defined as

$$(W)_{i,i} = \frac{w_i}{\sum_j a_{i,j}}. \quad (15)$$

In other words, the weighting scheme is a product of the statistical information and FPJ-weighting. Correspondingly the preconditioning matrix C should be redefined as

$$(C)_{j,j} = \frac{1}{\sum_i a_{i,j} w_i} \cdot \quad (16)$$

By these definitions, the component-wise reconstruction formula in Eqn. (7), with the replacement of matrix W_{FPJ} by matrix W, is

$$x_j^{(k+1)} = x_j^{(k)} + \frac{\sum_i a_{i,j} w_i \left(\frac{p_i - \sum_j a_{i,j} x_j^{(k)}}{\sum_j a_{i,j}} \right)}{\sum_i a_{i,j} w_i}. \quad (17)$$

The preconditioned gradient descent interpretation of weighted SART takes the component-wise form in Eqn. (13). As in the discussion of the convergence property of the conventional SART, the sufficient condition for convergence of weighted SART is the requirement that the spectral radius of the matrix $I - CA^T WA$ be less than 1.

It has been demonstrated that the weighting scheme does not have to be limited to statistical weighting scheme. One can design the weighting scheme such as the smooth window weighting scheme in helical cone beam imaging case to mitigate artefacts or to improve image quality due to axial truncation in circular cone imaging case [19]. We call this smooth window type weighting scheme as analytical weighting scheme to distinguish it from the statistical information and the forward projection model related weighting scheme. To simplify the notation, we incorporate this analytical weighting scheme into w_i, while keeping in mind that w_i can either be statistical weighting or analytical weighting or a multiplication of both. Because our definition of w_i is extended to including designed weights, it can be easily seen that Eqn. (17) reduces to the conventional SART when weights are constant for all projection data.

If we adopt the following definitions of the diagonal weighting matrix W and the preconditioning matrix C,

$$(W)_{i,i} = w_i \quad (18)$$

and

$$(C)_{j,j} = \frac{1}{\sum_i (a_{i,j} w_i \sum_k a_{i,k})}, \quad (19)$$

the component-wise reconstruction formula by Eqn. (7) is

$$x_j^{(k+1)} = x_j^{(k)} + \frac{\sum_i a_{i,j} w_i \left(p_i - \sum_j a_{i,j} x_j^{(k)} \right)}{\sum_i (a_{i,j} w_i \sum_k a_{i,k})}. \quad (20)$$

One recognize that Eqn. (20) is the formula in [17] without the regularization term. It should be noted that in this alternative derivation, we do not have to seek any surrogate as was originally proposed in [17]. Another point we want to make is that Eqn. (20) in general cannot reduce to Eqn. (12) if we insist that the weighting scheme be the statistical information, because statistical information in general are not uniform. The preconditioned gradient descent interpretation of Eqn. (20) can be easily obtained by simply adding a relaxation parameter in front of the image updating term. We emphasize that the weighting scheme w_i in Eqn. (20) can also be designed to improve image quality in particular imaging problems.

IV. Summary

In this paper, we interpreted two variants of the conventional SART, namely, weighted versions of the SART, following the interpretation of the conventional SART [11-13]. As a by-product, we proposed an alternative derivation of the WLS variant of SART without seeking any surrogate and provided a unified interpretation of it. The analysis in this paper provided some theoretic foundation to our proposed weighted SART in our previous work [16].

Acknowledgment

The author thanks Dr. Zamyatin at TMRU for numerous fruitful discussion of the project.

References

[1] G. N. Hounsfield, A method and apparatus for examination of a body by radiation such as X or Gamma radiation, Patent Specification 1283915, Longdon, England (1972).

[2] R. Gordon, R. Bender and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography", *J. Theor. Biol.* Vol. 29, pp. 471-481, 1970.

[3] S. Kaczmarz, "Angenaherte auflosung von system linearer gleichungen", *Bull. Acad. Polon. Sci. Lett. A*, Vol. 35, pp. 335-357, 1937.

[4] A. C. Kak and M. Slaney, "*Principles of Computerized Tomographic Imaging*", IEEE Press. 1988.

[5] E. Sidky, C. Kao and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-ray Sci. & Tech.*, vol. 14, pp. 119-139, 2006.

[6] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-2, 1982, pp. 113-122.

[7] Z. P. Liang and P. C. Lauterbur, Principles of Magnetic Resonance Imaing. Piscataway, NJ: IEEE Press, 2000.

[8] C. Riddell, "Parallel ART," *The CT Meeting*, 2010, pp. 128-131.

[9] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, 1984, pp. 81-94.

[10] G. Wang and M. Jiang, "Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART)," *J. of x-ray Sci. & Tech.*, vol. 12, 2004, pp. 169-177.

[11] M. Jiang and G. Wang, "Convergence studies on iterative algorithms for image reconstruction," *IEEE Trans. Med. Imag.*, Vol. 22, pp. 569-579, 2003.

[12] M. Jiang and G. Wang, "Convergence of simultaneous algebraic reconstruction technique (SART)", *IEEE Trans. Imag. Proc.*, Vol. 12, pp. 957- 961, 2003.

[13] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT", *IEEE Trans. Med. Imag.*, Vol. 27, pp. 918-924, 2008.

[14] J. Thibault, K. Sauer, C. Bouman and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, 2007, pp. 4526-4544.

[15] T. Kohler, R. Proksa and T. Nielsen, "SNR-weighted ART applied to transmission tomography", *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4, 2003, pp. 2739–2742.

[16] D. Shi, Y. Zou and A. Zamyatin, "Weighted Simultaneous Algebraic Reconstruction Technique," *Fully 3D*, pp. 157, 2011.

[17] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography", *IEEE Trans. Med. Imag.*, Vol. 21, pp. 89-99, 2002.

[18] R. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, 1985, pp. 252-255.

[19] A. Zamyatin, D. Shi and M. Dinu, "Extension of Axial Coverage and Artifact Reduction in Iterative Reconstruction in Computed Tomography," *IEEE MIC 2011*, MIC21.S-102.

Connecting image sparsity and sampling in iterative reconstruction for limited angle X-ray CT

Jakob S. Jørgensen*, Emil Y. Sidky† and Xiaochuan Pan†
*Department of Applied Mathematics and Computer Science, Technical University of Denmark
Matematiktorvet, bygning 303B, 2800 Kgs. Lyngby, Denmark. Email: jakj@imm.dtu.dk
†Department of Radiology, The University of Chicago
5841 S. Maryland Avenue, Chicago, IL 60637, USA. Email: {sidky,xpan}@uchicago.edu

***Abstract*—A possible quantitative relation between the image sparsity and the number of CT projections views sufficient for accurate reconstruction through 1-norm minimization is investigated empirically. In the setting of full and limited angular range fan-beam and circular cone-beam CT the average number of sufficient views is determined as function of phantom image sparsity over ensembles of randomly generated phantom images. For two phantom classes with different degrees of structure we find a quantitative relation as well as a sharp transition from inaccurate to accurate solution.**

I. Introduction

In the past few years, sparsity-exploiting image reconstruction methods for low-dose computed tomography (CT) have gained interest motivated by the field of compressed sensing (CS) [1], [2]. Numerous studies have demonstrated the potential for accurate reconstruction from a reduced number of measurements both in simulation and on clinical data. As the initial proof-of-concept has been carried out, the excitement over potential large data reduction is developing into new questions on what is missing before these techniques become standard practice [3].

Many factors affect reconstruction quality of sparsity-exploiting methods, including the amount and quality of data, the choice of algorithm and underlying optimization problem and the accuracy with which it is solved as well as the complexity of the test phantom – the topic of the present study. Typically, sparsity-exploiting methods involve many parameters that must be set in just the right way to get a favorable reconstruction, and the large size of realistic CT problems make exhaustive parameter space exploration infeasible. As a result, reconstruction quality of sparsity-exploiting methods remains less understood than for analytical methods.

Recently, we have been studying the role of phantom image complexity for reconstruction quality. Specifically, we have been quantifying the amount of undersampling to expect of a CS-based method in CT [4] and assessing the role of the image sparsity [5], i.e., the number of nonzero pixel values. Image sparsity is a key concept in CS but has to our knowledge not been addressed systematically in CT. In [5], we developed a so-called relative sampling-sparsity (RSS) diagram for investigating a connection between the image sparsity and the number of CT projection views required for accurate reconstruction in the setting of few-view, full angular range CT. The purpose of the present paper is two-fold: to extend the approach to study limited angle problems and to verify the connection between image sparsity and sufficient sampling predicted by small-scale 2-D fan-beam simulations on a 3-D circular cone-beam case.

II. Materials and methods

A. Sparsity-exploiting image reconstruction methods

Sparsity-exploiting methods are motivated by CS-results demonstrating that an image can be reconstructed accurately from a reduced number of measurements [2]. The assumption is that the image is sparse, that is, has a representation with few nonzero coefficients, for example pixel values. For certain discretized forward operators such as partial Fourier matrices and matrices with elements drawn from a Gaussian distribution, theoretical results state how many measurements are needed for *guaranteed* accurate reconstruction of an image of a given sparsity. For system matrices in CT, however, we are unaware of such guarantees, but can investigate a possible connection between sparsity and the number of measurements needed for accurate reconstruction empirically. The establishment of such a connection will provide insight into the amount of undersampling to expect for images of given a given sparsity.

Based on the so-called phase diagram introduced by Donoho and Tanner [6], we proposed in [5] specific for X-ray CT the relative sparsity-sampling (RSS) diagram for studying this connection empirically. Using the diagram we demonstrated the existence of a sharp transition from inaccurate to accurate reconstruction as function of the sparsity and number of measurements for X-ray CT with a 2-D few-view full-angular range scanner configuration. In the present work, we study a limited-angle case using the RSS-diagram.

B. Scanner configuration

We consider a 2-D fan-beam scanner configuration with N_v projections equi-distributed over 360° (full angular-range) or 90° (limited angular-range). The image is restricted to a disk-shaped mask within a $N_\mathrm{side} \times N_\mathrm{side}$ square image, which makes the number of pixels approximately $N = \pi/4 \cdot N_\mathrm{side}^2$. The source-to-center distance is set to $2N_\mathrm{side}$ and the fan-angle to 28.07° for illuminating the disk-shaped image. The detector consists of $N_\mathrm{b} = 2N_\mathrm{side}$ bins, which makes the total number of measurements $M = 2N_\mathrm{side}N_\mathrm{v}$. The line-intersection method

is used for computing X-ray path lengths through the image pixels, each ray yielding an equation of the form

$$b_i = \sum_{j=1}^{N} A_{ij} x_j, \qquad i = 1, \ldots, M, \tag{1}$$

where A_{ij} is the path length of the ith ray through the jth pixel and the system matrix A is of size $M \times N$.

We also consider a 3-D circular cone-beam scanner configuration with circular source trajectory over the same angular ranges. The object is then restricted to a ball-shaped mask within a $N_{\text{side}} \times N_{\text{side}} \times N_{\text{side}}$ cube image and each projection has size $2N_{\text{side}} \times 2N_{\text{side}}$ detector bins.

C. Phantom classes

We use the class of phantoms introduced in [5] called the p**-power** class. The class is originally described in [7] as a background breast tissue model, here followed by thresholding to create zero-valued pixels for obtaining sparse images suited for the experimental design of the present study. The parameter p governs the amount of structure. We can generate random instances of a desired target sparsity from the p**-power** phantom class and refer to a set of such instances as an ensemble. In the present study we consider $p = 0$ and $p = 2$; examples of phantom instances are seen in Fig. 1. The reason for using different phantom classes is to see if sparsity alone can explain the sampling needed for accurate reconstruction or other factors, here structure, play a role as well.

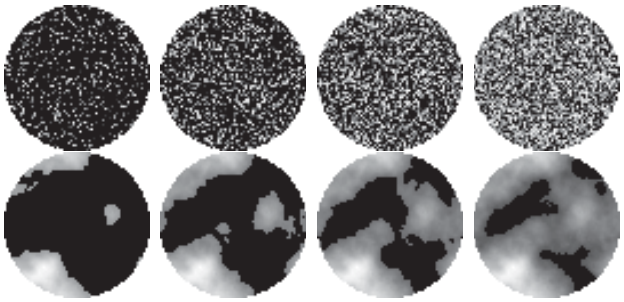


Fig. 1. p**-power** phantom instances. Top, bottom: Structure parameter $p = 0, 2$. Left to right: relative sparsity $\kappa = 0.2, 0.4, 0.6, 0.8$. Gray scale: [0,1].

D. Reconstruction problems and algorithms

For reconstruction, we consider the optimization problem

$$\mathbf{L_1}: \quad x_{\mathbf{L_1}} = \operatorname{argmin} \|x\|_1 \tag{2}$$

$$\text{s.t.} \quad Ax = b. \tag{3}$$

We wish to solve the optimization problem very accurately to avoid false conclusions based on inaccurate solutions. For this purpose, we employ the general-purpose commercial optimization software MOSEK [8], which uses a state-of-the-art primal-dual interior-point method. $\mathbf{L_1}$ can be recast as a linear program (LP), a standard optimization problem to which MOSEK produces a certified primal-dual solution.

For faster solution of the large-scale 3-D problem, we solve instead the problem

$$\mathbf{L_1^\delta}: \quad x_{\mathbf{L_1^\delta}} = \operatorname{argmin} \|x\|_1 \tag{4}$$

$$\text{s.t.} \quad \|Ax - b\|_2^2 \le \delta^2 \tag{5}$$

where the scalar parameter δ acts as a regularization parameter governing the size of the allowed data misfit. For small values of δ and consistent data, the $\mathbf{L_1^\delta}$ solution closely approximates the $\mathbf{L_1}$ solution. We use the Chambolle-Pock algorithm 1 described in [9] with $\delta = 10^{-5}$.

E. Simulation set-up

We create a phantom instance x_{orig} with $N_{\text{side}} = 64$ from one of the p**-power** classes and compute the ideal data $b = Ax_{\text{orig}}$ using different numbers of views, $N_{\text{v}} = 2, 4, 6, \ldots, 32$. We reconstruct by solving $\mathbf{L_1}$ to obtain $x_{\mathbf{L_1}}$. Reconstruction error is measured as the relative 2-norm error to the original, $\|x_{\mathbf{L_1}} - x_{\text{orig}}\|_2 / \|x_{\text{orig}}\|_2$. We accept $x_{\mathbf{L_1}}$ as perfectly recovering x_{orig} if the error is below a threshold of $\epsilon = 10^{-4}$.

With the chosen scanner configuration, we find for both 360° and 90° data that at $N_{\text{v}}^{\text{suf}} = 26$ or more views the system matrix has full column rank, causing x_{orig} to be the unique solution to $Ax = b$. At fewer views, the linear system is underdetermined, with infinitely many solutions and 1-norm minimization is used for selecting a sparse solution. Using $N_{\text{v}}^{\text{suf}}$ as a reference point of having sufficient—or full—sampling, we call $\mu = N_{\text{v}}/N_{\text{v}}^{\text{suf}}$ the relative sampling.

III. Results

A. 2-D fan-beam simulation results: Single phantom instances

First, we wish to demonstrate that $\mathbf{L_1}$ can perfectly recover the original image from 90° data, very similar to what we observed in [5] for 360° data. Fig. 2 shows reconstructions for both 360° and 90° data for $N_{\text{v}} = 6, 8, 10, 12$ of a 0**-power** phantom instance (no structure) and relative sparsity $\kappa = 0.2$. Also shown are difference images with the original to better visualize the transition to recovery. In both cases, we see that at $N_{\text{v}} = 12$ the reconstruction is numerically exact, as the difference images consist only of zeros. Interestingly, $\mathbf{L_1}$ reconstruction of a 0**-power** instance does not appear to be more difficult with the limited angular range of 90°.

We repeat the same experiment with a 2**-power** phantom instance of more structure and show results in Fig. 3. In this case, $N_{\text{v}} = 10$ suffices for accurate reconstruction from the 360° data, while $N_{\text{v}} = 12$ is needed for the 90° data. Apparently, from 360° data the structured phantom is easier to reconstruct than the unstructured, while from 90° data no difference due to structure is seen.

We repeat the experiment for relative sparsity of the 0**-power** phantom instance increased from $\kappa = 0.2$ to 0.4, 0.6 and 0.8. In Fig. 4, reconstruction errors from 360° data are plotted against numbers of views for the four κ-values. The jump to an accurate solution at $N_{\text{v}} = 12$ for $\kappa = 0.2$ from Fig. 2 is recognized. Similar jumps at $N_{\text{v}} = 16, 20, 24$ occur for $\kappa = 0.4, 0.6, 0.8$, and we conclude that the number of

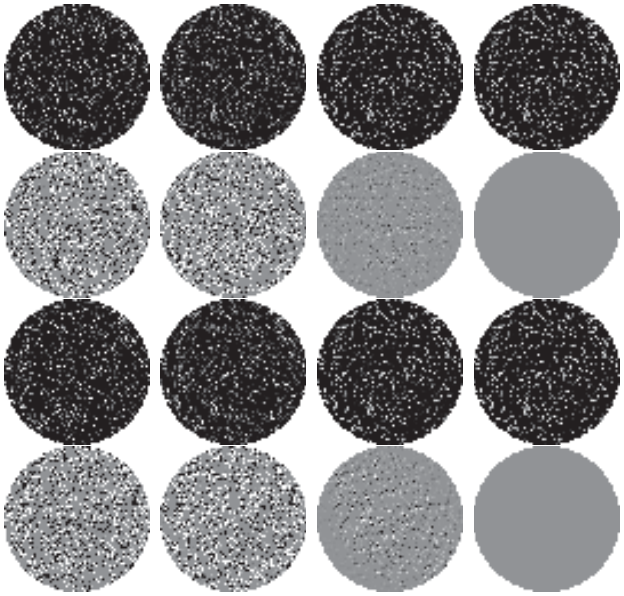


Fig. 2. Left to right: Reconstructions from $N_{\rm v} = 6, 8, 10, 12$ views of a 0-**power** phantom instance of relative sparsity $\kappa = 0.2$. 1st/3rd row: $360°/90°$ data reconstructions. Gray scale: $[0, 1]$. 2nd/4th row: $360°/90°$ data reconstructions minus original image. Gray scale: $[-0.1, 0.1]$.

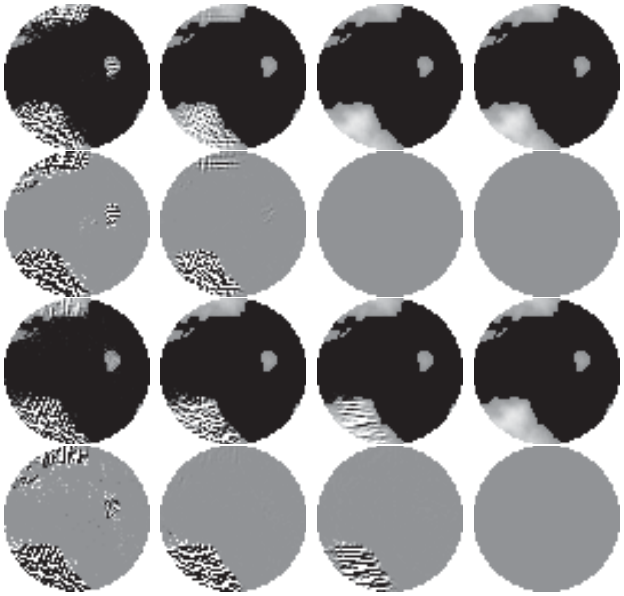


Fig. 3. Same as Fig. 2 for 2-**power** instance of relative sparsity $\kappa = 0.2$.

views needed for accurate $\mathbf{L_1}$-reconstruction appears to grow in a simple way with the relative sparsity κ. Put in another way, we see that images with fewer nonzero pixels admit a larger undersampling relative to the full-sampling reference point of $N_{\rm v}^{\rm suf} = 26$, as marked by the vertical line in Fig. 4.

B. RSS-diagrams: Multiple phantom instances

A natural question at this point is whether these observations are general or depend on the particular phantom instances used in Figs. 2 and 3. To answer the question, we repeat the experiment for 100 different phantom instances at each

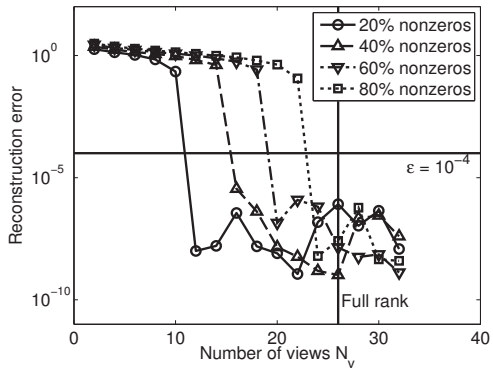


Fig. 4. Reconstruction errors $\|x_{\mathbf{L_1}} - x_{\rm orig}\|_2 / \|x_{\rm orig}\|_2$ as function of numbers of views $N_{\rm v}$ for relative sparsity values $\kappa = 0.2, 0.4, 0.6, 0.8$. In all cases, a steep jump from inaccurate to accurate solution is seen and the $N_{\rm v}$ at which the jump occurs increases with relative sparsity. The vertical line marks the lowest $N_{\rm v}$ at which the system matrix has full rank.

of the relative sparsity values $\kappa = 0.2, 0.4, 0.6, 0.8$. At each $N_{\rm v} = 2, 4, 6, \ldots, 32$ we record the percentage of phantom instances that are reconstructed to within a reconstruction error of $\epsilon = 10^{-4}$. The resulting percentages for the 0-**power** and 2-**power** phantom classes and $360°$ and $90°$ data are shown in what we call RSS-diagrams in Fig. 5. Each rectangle represents the percentage of phantoms recovered, ranging from 0% (black) to 100% (white) and shown as function of relative sparsity κ and relative sampling μ. For example, the black bottom left rectangle corresponds to $\kappa = 0.2$ and 2 views, i.e., $\mu = 2/26 \approx 0.08$. In all four cases we recognize the simple connection between relative sparsity and relative sampling sufficient for accurate reconstruction. For the 0-**power** class we observe a very sharp transition from inaccurate to accurate reconstruction in the sense that almost no difference in the relative sampling needed for accurate reconstruction exists among the 100 phantom instances. Furthermore, the RSS-diagrams for $360°$ and $90°$ data are identical, which supports our earlier conclusion that $\mathbf{L_1}$-reconstruction of the 0-**power** phantom class is unaffected by the limited angular range.

For the 2-**power** class, the transition from inaccurate to accurate reconstruction is slightly more gradual and for the $360°$ data occurs about one rectangle (2 views) lower than for the 0-**power** class as well as for the 2-**power** class with $90°$ data. We conclude that for the more structured phantom class 2-**power**, the limited angular range does make accurate reconstruction with $\mathbf{L_1}$ more difficult.

C. 3-D circular cone-beam simulation results

A practical use of the observed connection between relative sparsity and the relative sampling required for accurate reconstruction is to predict how many views will be needed in other and more difficult-to-simulate scenarios. In [5] we showed that the RSS-diagrams are essentially independent of the image size $N_{\rm side}$, so that we can predict sufficient numbers of views at larger pixel arrays based on RSS-diagrams from smaller pixel arrays such as 64×64. Here, we consider predicting the sufficient number of views on a different but related scanner

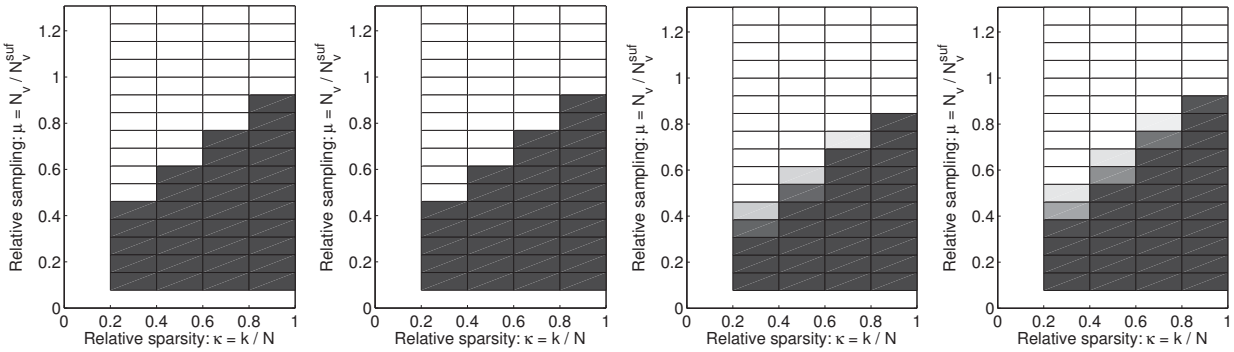


Fig. 5. RSS diagrams: Percentage of accurately reconstructed phantom images as function of relative sparsity and relative sampling. Black=0%, white=100%. Left to right: 0-**power** class with 360° data, 0-**power** class with 90° data, 2-**power** class with 360° data, 2-**power** class with 90° data.

configuration, namely 3-D circular cone-beam. We use a 3-D phantom instance of the 2-**power** class and size $N_{\text{side}} = 32$ with relative sparsity $\kappa = 0.2$. Using the N_{side}-independence of the RSS-diagram we expect at the fifth rectangle from below in the $\kappa = 0.2$ column, which for $N_{\text{side}} = 32$ corresponds to $N_{\text{v}} = 5$, to see a difference between 360° and 90° data. Selected slices of the 3-D $\mathbf{L}_\mathbf{1}^\delta$-reconstructions are shown in Fig. 6 and show excellent agreement with the expectation, as the 360° reconstruction is accurate while the 90° one is not. Interestingly, the central slice, which corresponds precisely to the previous 2-D CT configuration, appears to contradict our expectation as accurate reconstruction is observed in both cases. We explain this by the large degree of sparsity in this plane of the particular phantom instance, because other planes in the 90° reconstruction show prominent errors.

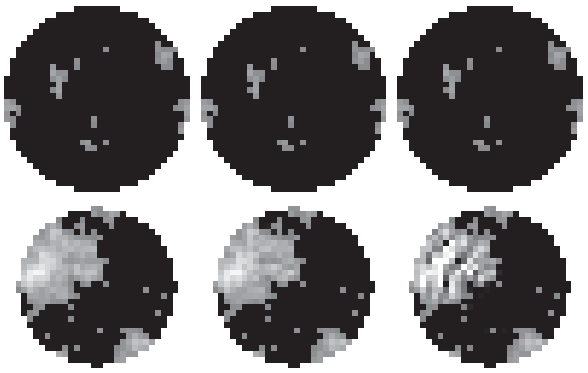


Fig. 6. Top row, left: Central slice (17 of 32 slices, parallel to the plane of the source trajectory) of the $32\times32\times32$ phantom instance from the 2-**power** class of relative sparsity $\kappa = 0.2$. Middle: same slice of 3-D reconstruction from 360° data. Right: same slice of 3-D reconstruction from 90° data. Bottom row: Same for off-central slice (8 of 32). Gray scale: $[0, 1]$.

IV. Discussion and conclusion

The results presented here demonstrate empirically a relation between sparsity of the image to be reconstructed and the average number of fan-beam views required for accurate reconstruction with $\mathbf{L_1}$, both on full angular range and 90° limited angular data. Structured phantoms were found to be accurately reconstructed from slightly fewer views than unstructured phantoms of same sparsity, indicating that image sparsity can only explain some of the variation of the required number of views. The relation found can be used for understanding what undersampling levels to expect when reconstructing sparse images. The RSS-diagram can serve as a tool to investigate such a relation on other sparsity-exploiting methods, e.g., total variation for image gradient sparsity.

Acknowledgment

This work was supported in part by the project CSI: Computational Science in Imaging (The Danish Research Council for Technology and Production Sciences grant 274-07-0065), and in part by The Danish Ministry of Science, Innovation and Higher Education's Elite Research Scholarship. This work was supported in part by NIH R01 grants CA158446, CA120540 and EB000225. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, 2006.

[3] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Prob.*, vol. 25, p. 123009, 2009.

[4] J. S. Jørgensen, E. Y. Sidky, and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT." *IEEE Trans. Med. Imag.*, vol. 32, pp. 460–473, 2013.

[5] J. S. Jørgensen, E. Y. Sidky, P. C. Hansen, and X. Pan, "Quantitative study of undersampled recoverability for sparse images in computed tomography," *Submitted*. [Online]. Available: http://arxiv.org/abs/1211.5658

[6] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing." *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, vol. 367, pp. 4273–4293, 2009.

[7] I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," *Med. Phys.*, vol. 37, pp. 1591–1600, 2010.

[8] MOSEK ApS, "MOSEK Optimization Software (www.mosek.com)," Copenhagen, Denmark. [Online]. Available: www.mosek.com

[9] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm," *Phys. Med. Biol.*, vol. 57, pp. 3065–3091, 2012.

Statistical-Analytic Regularized Reconstruction for X-ray CT

Kyle M.L. Champley and Harry E. Martz, Jr

***Abstract*—Limitations in X-ray flux lead to noisy CT images. One can sacrifice a small amount of contrast and resolution for a large gain in the signal-to-noise ratio of the CT image using nonlinear regularization techniques. Nonlinear regularized reconstruction requires computationally-expensive iterative techniques and thus efficient algorithms are needed to process the data.**

In this paper we use analytic X-ray CT image reconstruction techniques (e.g. filtered backprojection) to design preconditioners that improve the rate of convergence of the gradient descent algorithm. We show that our algorithms converge faster than gradient descent with a 2D ramp filter preconditioner using simulated conebeam helical X-ray CT data.

I. Introduction

Filtered backprojection (FBP) is a one-pass algorithm used to reconstruct X-ray CT images from the measured data. The algorithm is very well understood and provides great images in a short amount of time. In some cases the X-ray flux is limited. This may be due to dose limitations to the object being scanned, scan time, power of the X-ray source, or detector saturation. Low X-ray fluxes lead to noisy FBP reconstructions. Through regularized reconstruction we may reduce the amount of noise in the reconstructed image by sacrificing some contrast and resolution. Nonlinear regularization techniques such as total variation provide a generous tradeoff of improved signal-to-noise ratio for a moderate loss of contrast and resolution. Unfortunately these techniques require computationally-expensive iterative techniques to solve. With the help of advances in computing and new algorithms, iterative reconstruction is becoming the algorithm of choice in the case of noisy data. In addition to incorporating image priors into the reconstruction, iterative techniques allow one to take the statistics of the measurements into account.

One common choice is regularized weighted least squares. Here one attempts to minimize the following cost function

$$\Phi(f) := \frac{1}{2}(Pf - g)^T W(Pf - g) + S(f), \qquad (1)$$

where P is the digital X-ray transform (forward model of X-ray CT), f is a three-dimensional voxelized image, g is the measured data, W is a matrix of weights (often the inverse covariance matrix of g), and $S(\cdot)$ is a prior term that encourages smoothness of the solution, e.g., L^p norm of the magnitude of the gradient.

Many iterative algorithms designed to minimize convex cost functions take the form

$$f_{n+1} \quad = \quad f_n - \lambda_n d_n \qquad (2)$$

where $\lambda_n > 0$ is the step size and the descent direction, d_n, satisfies $d_n^T \Phi'(f_n) > 0$. The sequence $\{f_n\}$ converges to the minimizer of $\Phi(f)$. One such iterative algorithm is preconditioned gradient descent which is given by

$$\begin{aligned} f_{n+1} &= f_n - \lambda_n Q \Phi'(f_n) \\ \Phi'(f) &= P^T W(Pf - g) + S'(f), \end{aligned}$$

where $\Phi'(f)$ is the Jacobian (functional derivative) of $\Phi(f)$ and Q is positive definite. This paper is concerned with developing a preconditioner, Q, to improve the convergence speed of the above iteration.

II. Preconditioner Design

Preconditioners speed up the convergence rate by decreasing the condition number of the Hessian of the cost function, i.e.,

$$\text{cond}\left(Q\Phi''(f)\right) < \text{cond}\left(\Phi''(f)\right).$$

In our case, $\Phi''(f) = P^T W P + S''(f)$. Certainly $Q = [\Phi''(f)]^{-1}$ is an excellent choice for a preconditioner (and leads to the Newton-Raphson method), but finding this inverse is often more difficult than solving the actual problem. Thus we attempt to find an approximate left inverse of $\Phi''(f)$.

We develop our algorithm in continuous space. Thus we define $\mathcal{P}$ and $\mathcal{S}(\cdot)$ as the continuous analogs of P and $S(\cdot)$, respectively. We make no notational distinction between the continuous and discrete versions of f and g. Let $\mathcal{A}$ be a left inverse of $\mathcal{P}$, i.e., $\mathcal{A}\mathcal{P} = \mathcal{I}$, where $\mathcal{I}$ is the identity transform. In this paper, we mostly use $\mathcal{A}$ as FBP, or more accurately speaking, the discretization of $\mathcal{A}$ is FBP, but it can also be backprojection filtration, Katsevich algorithm [1], Tuy's algorithm [2], Finch's algorithm [3], or any other analytic inverse of the continuous X-ray transform. A left inverse does not exist for the discrete X-ray transform.

Lawrence Livermore National Laboratory, Livermore, CA 94550 (e-mail: champley1@llnl.gov).

Note that

$$\Phi''(f) \quad = \quad \mathcal{P}^*\mathcal{W}\mathcal{P} + \mathcal{S}''(f).$$

We make the coarse approximation, $\Phi''(f) \approx \mathcal{P}^*\mathcal{P}$. Since $\mathcal{A}$ is a left-inverse of $\mathcal{P}$, $\mathcal{P}\mathcal{A}$ is an orthogonal projection and thus

$$\begin{aligned} \mathcal{A}\mathcal{A}^*\Phi''(f) &\approx \mathcal{A}\mathcal{A}^*\mathcal{P}^*\mathcal{P} = \mathcal{A}(\mathcal{P}\mathcal{A})^*\mathcal{P} \\ &= \mathcal{A}\mathcal{P}\mathcal{A}\mathcal{P} = \mathcal{I}. \end{aligned}$$

Therefore, $\mathcal{A}\mathcal{A}^*$ is an approximate left inverse of the Hessian of the cost function and is a strong candidate for a preconditioner. We have also shown that when $\mathcal{A}$ is of filtered backprojection type, one can derive a backprojection filtration algorithm by $(\mathcal{A}\mathcal{A}^*)\mathcal{P}^*$. Now the preconditioned gradient descent step direction is given by

$$\begin{aligned} \mathcal{A}\mathcal{A}^*\Phi'(f) &= \mathcal{A}\mathcal{A}^*\left[\mathcal{P}^*\mathcal{W}(\mathcal{P}f - g) + \mathcal{S}'(f)\right] \\ &= \mathcal{A}\mathcal{W}(\mathcal{P}f - g) + \mathcal{A}\mathcal{A}^*\mathcal{S}'(f), \end{aligned}$$

since

$$\mathcal{A}\mathcal{A}^*\mathcal{P}^* = \mathcal{A}(\mathcal{P}\mathcal{A})^* = \mathcal{A}\mathcal{P}\mathcal{A} = \mathcal{A}.$$

One can find $\widehat{h}(\mathbf{x}, \boldsymbol{\xi})$ such that

$$\mathcal{A}\mathcal{A}^* f(\mathbf{x}) = \int_{\mathbb{R}^3} \widehat{f}(\boldsymbol{\xi})\widehat{h}(\mathbf{x}, \boldsymbol{\xi}) e^{2\pi i <\mathbf{x}, \boldsymbol{\xi}>} \, d\boldsymbol{\xi},$$

where $\mathbf{x} \in \mathbb{R}^3$ is the spatial coordinate of the image and $\boldsymbol{\xi} \in \mathbb{R}^3$ is the respective frequency in Fourier space. In the case of axial or helical acquisitions, one cannot use FFT operations to compute the above integral, and thus its computational complexity is $O(N^6)$. In the two-dimensional case,

$$\widehat{h}(\mathbf{x}, \boldsymbol{\xi}) \propto \sqrt{\xi_1^2 + \xi_2^2}.$$

In the three-dimensional case this is a fairly-accurate approximate relation.

III. Development of Algorithms

We investigate the performance of two algorithms motivated by the analysis in the previous section. The algorithms are given by equation (2), where

$$d_n \quad := \quad \mathcal{A}\mathcal{A}^*\left[\mathcal{P}^*\mathcal{W}(\mathcal{P}f_n - g) + \mathcal{S}'(f_n)\right], \qquad (3)$$

$$d_n \quad := \quad \mathcal{A}\mathcal{W}(\mathcal{P}f_n - g) + \frac{1}{2\pi N_{conj}}\mathcal{R}\mathcal{S}'(f_n), \qquad (4)$$

and $\mathcal{R}$ is the 2D ramp filter which, in arbitrary dimension, is given by

$$\mathcal{R}h(\mathbf{x}) := \int_{\mathbb{R}^n} \|2\pi\boldsymbol{\xi}\| \widehat{h}(\boldsymbol{\xi}) e^{2\pi i <\mathbf{x}, \boldsymbol{\xi}>} \, d\boldsymbol{\xi}$$

The term $N_{conj} \in \mathbb{Z}$ is equal to the mean number of conjugate rays. For axial data it is equal to one (short scan) or two (full scan) and for helical data is equal to $\lfloor 2/\widehat{\Delta} \rfloor$, where $\widehat{\Delta}$ is the normalized helical pitch. We will refer to these iterative algorithms as Statistical-Analytic Regularized Reconstruction 1 and 2 (SARR-1 and SARR-2), respectively. For either algorithm, λ_n is given by

$$\lambda_n = \frac{< \mathcal{P}d_n, \mathcal{W}(\mathcal{P}f_n - g) > + < d_n, \mathcal{S}'(f_n) >}{< \mathcal{P}d_n, \mathcal{W}\mathcal{P}d_n > + < d_n, \mathcal{S}''(f_n)d_n >}, \quad (5)$$

and is chosen to (approximately) minimize $\Phi(f_n - \lambda_n d_n)$.

Gradient descent only requires one forward ($\mathcal{P}$) and one backprojection ($\mathcal{P}^*$) per iteration, but SARR-1 requires two of each ($\mathcal{A}$ requires backprojection and $\mathcal{A}^*$ requires forward projection). Since $\mathcal{A}\mathcal{A}^*$ is merely a preconditioner, one does not need to implement a highly accurate forward/ back projector pair in the computation of $\mathcal{A}\mathcal{A}^*$. Thus one can mitigate the computational burden of these extra steps by using a coarse forward/ back projector pair in the computation of $\mathcal{A}\mathcal{A}^*$.

Regularized iterative filtered backprojection [4] is given by

$$f_{n+1} \quad = \quad f_n - \lambda\left[\mathcal{A}\mathcal{W}(\mathcal{P}f_n - g) + \mathcal{S}'(f_n)\right], \quad (6)$$

where $\lambda > 0$ is a constant step-size that is determined heuristically. This iteration does not converge to a minimizer of $\Phi(f)$ and the method to incorporate the regularization is also done by heuristic means. The SARR-2 algorithm is very similar to the regularized iterative filter backprojection algorithm, but SARR-2 is design to (approximately) minimize $\Phi(f)$ and the descent direction is approximately equal to the Newton-Rhapson descent direction. Note that since $\frac{1}{2\pi N_{conj}}\mathcal{R} \approx \mathcal{A}\mathcal{A}^*$, SARR-2 will not converge to the minimizer of $\Phi(f)$. We argue that the approximations made do not strongly effect the solution and thus it makes sense to choose the step size sequence λ_n by minimization of $\Phi(f_n - \lambda_n d_n)$. This sequence is given by equation (5), where d_n is given in equation (4). Here λ_n may be negative. Clearly $\Phi(f_{n+1}) \leq \Phi(f_n)$, but in most cases this algorithm will not converge to the unique minimizer of $\Phi(\cdot)$.

The 2D ramp filter has been studied as a preconditioner in many publications [5], [6], including attempts at approximating $(\mathcal{P}^T\mathcal{P})^{-1}$ as a non-stationary 2D ramp filter [7], but the authors do not know of any published work that uses $\mathcal{A}\mathcal{A}^*$ as a preconditioner. The closest algorithm to SARR-2 is the regularized iterative filtered back projection (RIFBP) algorithm. The calculation of the non-constant step sequence and the treatment of the regularization term in SARR-2 differs from RIFBP and is unique.

We let

$$\mathcal{S}(f) := \beta \int_{\mathbb{R}^n} \int_{S^{n-1}} h_{\delta,p}(\mathcal{D}_{\boldsymbol{\Theta}} f(\mathbf{x})) \, d\boldsymbol{\Theta} \, d\mathbf{x} \qquad (7)$$

$$h_{\delta,p}(t) := \begin{cases} |t|^p + \left(\frac{p}{2} - 1\right)\delta^p, & |t| > \delta, \\ \frac{p}{2}\delta^{p-2}t^2, & |t| \leq \delta, \end{cases} \qquad (8)$$

where $\mathcal{D}_{\boldsymbol{\Theta}}$ is the derivative in the $\boldsymbol{\Theta}$ direction, $p \geq 1$ and $\beta \geq 0$ controls the strength of the image prior term.

The requirement $p \geq 1$ is to ensure that $\mathcal{S}$ is convex. Note that $p = 1$ leads to total variation regularization and $p = 2$ leads to Tikhonov regularization.

Solving for the step sizes, λ_n, requires computation of $\mathcal{S}''(f)$. In practice we use a quadratic surrogate [8].

IV. Numerical Experiments

Helical cone-beam data is parameterized by

$$\mathcal{P}f(\alpha,\beta,v) := \int_{\mathbb{R}} f\left(\begin{bmatrix} R\cos\beta \\ R\sin\beta \\ \Delta\beta \end{bmatrix} + \frac{t}{\sqrt{1+v^2}}\begin{bmatrix} -\cos(\alpha+\beta) \\ -\sin(\alpha+\beta) \\ v \end{bmatrix}\right) dt.$$

When rebinned into cone-parallel coordinates it is given by

$$g(s,\varphi,v) := \mathcal{P}f(s,\varphi,v) = \int_{\mathbb{R}} f\left(\mathbf{y}(s,\varphi) + t\boldsymbol{\Theta}(\varphi,v)\right)\, dt$$

where

$$\begin{aligned}
\boldsymbol{\theta}(\varphi) &:= (\cos\varphi, \sin\varphi, 0)^T \\
\boldsymbol{\theta}^{\perp}(\varphi) &:= (-\sin\varphi, \cos\varphi, 0)^T \\
p(s,\varphi) &:= \Delta\left(\varphi - \sin^{-1}\left(\frac{s}{R}\right)\right) \\
\mathbf{y}(s,\varphi) &:= \sqrt{R^2-s^2}\boldsymbol{\theta}(\varphi) - s\boldsymbol{\theta}^{\perp}(\varphi) + p(s,\varphi)\widehat{\mathbf{z}} \\
\boldsymbol{\Theta}(\varphi,v) &:= \frac{1}{\sqrt{1+v^2}}\left[-\boldsymbol{\theta}(\varphi) + v\widehat{\boldsymbol{z}}\right]
\end{aligned}$$

for $s \in [s_0, s_{end}]$, $\varphi \in [\varphi_0, \varphi_{end}]$, and $|v| \leq v_{max}$.

TABLE I
Simulation and reconstruction settings.

Source to Iso-center Distance	320 mm
Source to Detector Distance	640 mm
Number of Detector Columns	320
Number of Detector Rows	64
Number of Views per Rotation	1000
Detector Element Size	2.09×2.09 mm^2
Table Feed	120 mm
Reconstruction Radius	160 mm
Reconstruction Height	64 mm
Cone Angle	$\pm 3^\circ$
Number of Voxels	$320 \times 320 \times 64$
p	1.1
δ	1.0 HU
ϵ	0.7

For the sake of our comparison we use the weighted filtered backprojection algorithm (WFBP) [9]. In continuous space the algorithm is given by

$$\begin{aligned}
\mathcal{A}g(\mathbf{x}) &:= \frac{1}{2\pi}\mathcal{P}_w^*\mathcal{R}\mathcal{C}g(\mathbf{x}), \\
\mathcal{C}g(s,\varphi,v) &:= \frac{1}{\sqrt{1+v^2}}g(s,\varphi,v) \\
\mathcal{P}_w^*g(x) &:= \int_{M(\mathbf{x})} g(-\mathbf{x}\cdot\boldsymbol{\theta}^{\perp},\varphi,v(\mathbf{x},\varphi))w(\mathbf{x},\varphi) \\
&\times \frac{1+v^2(\mathbf{x},\varphi)}{\sqrt{R^2-(\mathbf{x}\cdot\boldsymbol{\theta}^{\perp})^2}-\mathbf{x}\cdot\boldsymbol{\theta}}\, d\varphi \\
\mathcal{P}_w f(s,\varphi,\mathbf{x}) &= \mathcal{P}\{f(\mathbf{x})w(\mathbf{x},\varphi)\}(s,\varphi,v)
\end{aligned}$$

where $\mathcal{R}$ is the 1D ramp filter applied in the s coordinate and

$$\begin{aligned}
v(\mathbf{x},\varphi) &:= \frac{x_3 - \Delta\left(\varphi + \sin^{-1}\left(\frac{\mathbf{x}\cdot\boldsymbol{\theta}^{\perp}}{R}\right)\right)}{\sqrt{R^2-(\mathbf{x}\cdot\boldsymbol{\theta}^{\perp})^2}-\mathbf{x}\cdot\boldsymbol{\theta}} \\
w(\mathbf{x},\varphi) &:= \frac{\sqrt{R^2-(\mathbf{x}\cdot\boldsymbol{\theta}^{\perp})^2}-\mathbf{x}\cdot\boldsymbol{\theta}}{1+v^2(\mathbf{x},\varphi)} \\
&\times \frac{b_\epsilon\left(\frac{v(\varphi,\mathbf{x})}{v_{max}}\right)}{\sum_k b_\epsilon\left(\frac{v(\varphi+k\pi,\mathbf{x})}{v_{max}}\right)} \\
b_\epsilon(t) &:= \begin{cases} 1, & |t| \leq \epsilon, \\ \cos^2\left(\frac{\pi}{2}\frac{|t|-\epsilon}{1-\epsilon}\right), & \epsilon < |t| \leq 1, \\ 0, & |t| > 1, \end{cases} \\
M(\mathbf{x}) &:= \{\varphi \in [\varphi_0, \varphi_{end}] : |v(\varphi,\mathbf{x})| \leq v_{max}\}.
\end{aligned}$$

The original WFBP paper [9] does not include the $\mathcal{C}$ term, but we argue that its inclusion improves the accuracy of the algorithm. Note that the term

$$\frac{1+v^2(\mathbf{x},\varphi)}{\sqrt{R^2-(\mathbf{x}\cdot\boldsymbol{\theta}^{\perp})^2}-\mathbf{x}\cdot\boldsymbol{\theta}}$$

in the definition of $\mathcal{P}_w^*$ results from the divergent cone-parallel coordinate system. This term is automatically included in ray-driven, distance-driven [10], and separable footprint [11] backprojection algorithms, but not in voxel-driven backprojection. We reduce some of the computational burden combining the filtering operations of $\mathcal{A}$ and $\mathcal{A}^*$

$$\mathcal{A}\mathcal{A}^* = \frac{1}{(2\pi)^2}\mathcal{P}_w^*\mathcal{R}\mathcal{C}\mathcal{C}\mathcal{R}\mathcal{P}_w = -\frac{1}{(2\pi)^2}\mathcal{P}_w^*\mathcal{D}^2\mathcal{C}^2\mathcal{P}_w,$$

where $\mathcal{D}^2 = \frac{\partial^2}{\partial s^2}$.

We implement the forward/ back projection pair (P, P^*) by the Separable Footprint method [11]. We used the Shepp-Logan digital ramp filter which is given

$$h_{sl}[n] = \frac{1}{\pi\left(\frac{1}{4}-n^2\right)}.$$

The directional derivatives in $S(f)$ were computed using forward differences on the 26 nearest neighbors. The statistical weighting matrix, W, was given by $W = \text{diag}(e^{-g})$. The CT data was simulated using analytic ray-tracing methods and Poisson noise was added to the transmission data (pre-log corrected). Parameters for the simulation and reconstruction are listed in table I. The attenuating object is composed ellipsoids of varying attenuation within a cylinder of water and is shown in figure 1. The edges of the two largest ellipsoids are apodized by a window given by $\sqrt{1-\|\mathbf{x}\|^2}$ to avoid excessive artifacts in the reconstruction.

We test our algorithms versus gradient descent with the ramp filter as a preconditioner. All iterations start

TABLE II
SIGNAL-TO-NOISE RATIO (SNR) AND RELATIVE RMSE.

	SNR	relative RMSE
WFBP	219	0.66%
PGD	313	0.54%
SARR-1	267	0.56%
SARR-2	453	0.29%

with $f_0 = Ag$. We ran five iterations of the preconditioned (2D ramp filter) gradient descent (PGD), SARR-1, and SARR-2 algorithms.

V. RESULTS

The central axial slice of the reconstructed images with the WFBP, PGD, and SARR-2 algorithms are shown in figure 1. The signal-to-noise ratio and relative root mean square error (RMSE) (RMSE is normalized by $\sqrt{\sum_i f_{true,i}^2}$, where f_{true} is the true CT image) of each algorithm is shown in table II. We do not show the results for SARR-1 because they are very similar to PGD, only slightly noiser. The extra forward and backprojection operations act as a low pass filter on the step direction and thus are unable to correct high frequency artifacts as efficiently.

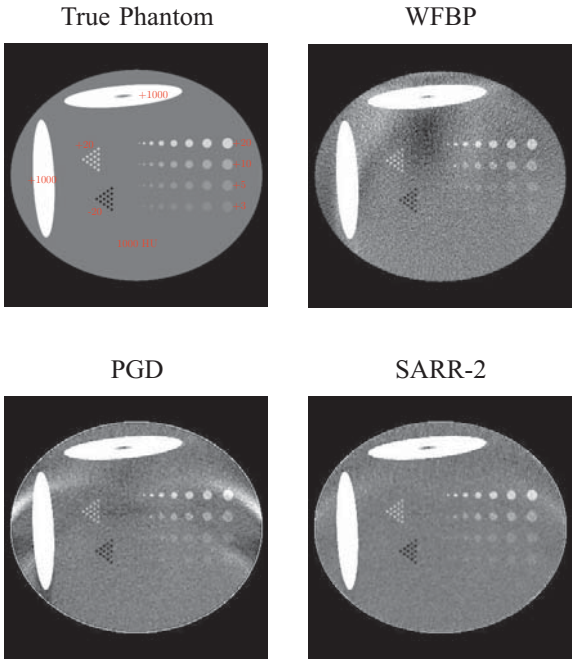


Fig. 1. Central axial slice of CT images. Window level [980, 1030] HU.

None of the algorithms converged in five iterations.

VI. CONCLUSION AND DISCUSSION

We used analytic X-ray CT image reconstruction techniques (e.g. filtered back projection) to develop fast iterative techniques (SARR-1 and SARR-2) to find the regularized weighted least squares solution. Our experiments show that the SARR-2 algorithm produces images with superior image quality (highest SNR, lowest RMSE, and least artifacts) compared to WFBP, PGD, and SARR-1 after five iterations.

One of the drawbacks of the methods presented in this paper is the efficient computation of $\mathcal{A}\mathcal{A}^*$. In SARR-1 we proposed speeding up this computation by using a fast forward / back projector pair (such as a ray-driven method [10]). Other than extra computation this also blurred the step direction too much. Computation of $\mathcal{A}\mathcal{A}^*$ was completely avoided in the SARR-2 algorithm by using accurate simplifications. Our results show that these approximations do not significantly effect the end result.

VII. ACKNOWLEDGEMENT

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

REFERENCES

[1] Alexander Katsevich. Analysis of an exact inversion algorithm for spiral cone-beam CT. *Phys. Med. Biol.*, 47:2583–2597, 2002.
[2] Heang K. Tuy. An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math*, 43:546–552, 1983.
[3] David Finch. Cone beam reconstruction with sources on a curve. *SIAM J. Appl. Math*, 45:665–673, 1985.
[4] Johan Sunnegardh. Iterative filtered backprojection methods for helical cone-beam CT. *PhD Dissertation, Linkoping Studies in Science and Technology*, pages 1–168, 2009.
[5] Jeffrey A. Fessler and Scott D. Booth. Conjugate-gradient preconditioning methods for shift-variant pet image reconstruction. *IEEE Trans. Imag. Proc.*, 8:688–699, 1999.
[6] Lin Fu, Bruno DeMan, Kai Zeng, Thomas M. Benson, Zhou Yu, Guangzhi Cao, and Jean-Baptiste Thibault. A preliminary investigation of 3d preconditioned conjugate gradient reconstruction for cone-beam ct. *Proc. of SPIE*, 8313:1–9, 2012.
[7] Neal H. Clinthorne, Tin-Su Pan, Ping-Chun Chiao, W. Leslie Rogers, and John A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imag.*, 12:78–83, 1993.
[8] Zhou Yu, Jean-Baptiste Thibault, Ken Sauer, Charles Bouman, and Jiang Hsieh. Accelerated line search for coordinate descent optimization. *IEEE Nuclear Science Symposium Conference Record*, pages 2841–2844, 2006.
[9] Karl Stierstorfer, Annabella Rauscher, Jan Boese, Herbert Bruder, Stefan Schaller, and Thomas Flohr. Weighted FBP-a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch. *Physics In Medicine and Biology*, 49:2209–2218, 2004.
[10] Bruno De Man and Samit Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49:2463–2475, 2004.
[11] Yong Long, Jeffrey A. Fessler, and James M. Balter. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29:1839–1850, 2010.

Few-View Image Reconstruction with Fractional-Order Total Variation

Yi Zhang*, Wei-Hua Zhang*, Yi-Fei Pu*, Yin-Jie Lei†, Hu Chen*, Meng-Long Yang‡ and Ji-Liu Zhou*
*College of Computer Science
Sichuan University, Chengdu 610065, Sichuan, China
†School of Computer Science and Software Engineering
The University of Western Australia, Crawley WA 6009, Perth, Australia
‡The College of Aerospace Science and Engineering
Sichuan University, Chengdu 610065, Sichuan, China
Corresponding author's email: yzhang@scu.edu.cn

***Abstract*—This work presents a novel Computed Tomography (CT) reconstruction method for few-view problem based on fractional calculus. To overcome the disadvantages of total variation minimization method, we apply fractional-order total variation in our method instead of traditional total variation and the numerical scheme for our method is also given. We use the root mean square error (RMSE) as a referee. The numerical experiments demonstrate that our method achieves better performance than existing reconstruction methods, including filtered back projection (FBP), expectation maximization (EM) and total variation with projection on convex sets (TV-POCS).**

I. Introduction

At present, Computed Tomography (CT) is still widely used in clinical diagnosis and industrial inspection. It is well known that X-ray is harmful to human body and it may cause genetic diseases [1]. Reducing the radiation dose is playing a more and more important role in our medical imaging topics. To overcome this problem, there are many different methods which can be divided into two categories: the first one is to lower the X-ray flux and the second one is to reduce the X-ray numbers across the human body. The former will lead to noisy projection data. The later will cause incomplete projection data in the forms of few-view, limited-angle, interior CT, etc. Although both categories of methods can reduce the radiation dose effectively, in this paper, we only focus on few-view problem to validate the proposed model and it is direct to extend our method to other topics.

Conventional analytic methods such as filtered back projection (FBP) require a high sampling rate which can lead to a satisfactory image quality. Incomplete projection data will cause bad visual effects. Iterative algorithms are used to deal with this problem. Over the past decades, much effort was spent on this problem and lots of methods were proposed such as algebraic reconstruction technique (ART) [2], simultaneous algebraic reconstruction technique (SART) [3], expectation maximization (EM) [4], etc. However, when the projection data are highly undersampling, without extra prior information, it is very hard to converge to the correct solution. Compressive sensing (CS) is one of the most popular methods developed to handle an under-determined problem [5]. Inspired by these works, Sidky introduced total variation (TV) minimization to incomplete projection data reconstruction and obtained very good result [6]. But TV is based on a famous assumption that the signal is piecewise smooth. This assumption makes TV algorithm suffer from over-smoothing which means TV can not perfectly preserve structure information like edges and shapes. To alleviate this side-effect of TV, many methods were proposed. To fix the over-smoothing and staircase effect, You and Kaveh used Laplacian of image to replace the first-order gradient in TV model [7]. Chan et al, added a nonlinear fourth-order diffusive term to the Euler-Lagrange equations of the variational TV model [8]. Lysaker et al, gave two high-order functionals to measure oscillations in noisy images [9]. All these methods can relieve the side effects of TV to certain extent, but results in another problem, such as speckle noise.

Recently, as a new mathematical tool, fractional calculus has been used in image processing. Zhang and Wei constructed fractional bounded variation (fBV) space to recover more texture information from noisy images [10]. Bai and Feng generalized the anisotropic diffusion model into a fractional-order version and when the order is 1.8 and 2.2, the peak signal to noise ratio (PSNR) reached the maximum [11]. Two fractional-order image inpainting models which respectively correspond with image domain and wavelet domain were given by Zhang et al [12]. Zhang et al, first brought fractional calculus into medical imaging issues. They proposed two different fractional-order models to suppress metal artifacts in CT imaging [13][14]. In these results, fractional calculus showed some special strengths, such as multi-scale, robustness and nonlinearity. Especially in [10] and [13], fractional-order total variation was shown that it can effectively mitigate the over-smoothing effect without introducing other drawbacks like high-order methods [7][8].

In this paper, we will introduce fractional calculus to solve the few-view problem in the form of fractional-order total variation which is calculated from fractional-order gradient of images. Unlike conventional image processing techniques which usually handle an image pixel by pixel, the proposed method process the image patch by patch. These patches are extracted with our proposed fractional-order masks [15].

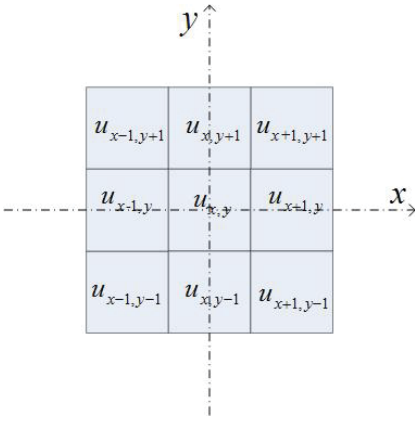


Fig. 1: Position illustration for $u_{x,y}$

.

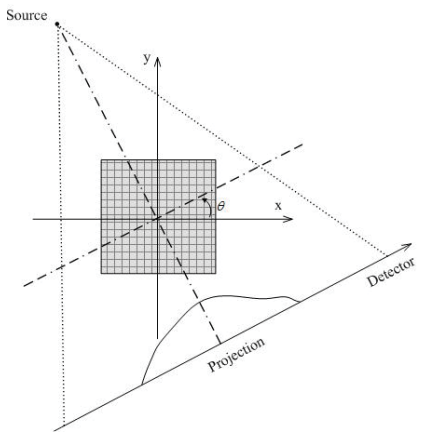


Fig. 2: Fan-beam CT geometry configuration

.

The theoretical details are described in section 2 and some numerical experiments will be provided in section 3. Finally, the conclusion will be given.

II. Method

A. The TV-based image reconstruction method

Our method is an extended version of TV-based image reconstruction. In this section, we will first give a brief description of this method. Given a 2-dimensional image $u = u_{x,y}$, whose size is $M \times N$, $x \in [1, M]$ and $y \in [1, N]$. For any u, the gradient operator is defined as

$$\nabla u = (\Delta_x u, \Delta_y u), \tag{1}$$

where Δ_x and Δ_y are the first-order differential operators along x-axis and y-axis respectively. Δ_x and Δ_y can be represented as

$$\begin{aligned} \Delta_x u &= u_{x,y} - u_{x-1,y}, \\ \Delta_y u &= u_{x,y} - u_{x,y-1}. \end{aligned} \tag{2}$$

An illustration is given as Fig.1.

In traditional CT imaging problem, the sampling procedure can be seen as a discrete linear transform,

$$Au = f, \tag{3}$$

where A is the system matrix which is comprised of I row vectors and $f = (f_1, f_2, \ldots, f_I)^T$ is the measurement vector . The individual elements of the system matrix are A_{ij} and $j = 1, 2, ..., J$. Without losing generality, the fan-beam projection geometry can be demonstrated in Fig. 2.

To solve the linear system in (3), the TV-based image reconstruction algorithm which was used to deal with the few-view limitation is to optimize the following problem [6]:

$$\min \|u\|_{\mathrm{TV}} \quad \text{subject to } u \geq 0,\ Au = f, \tag{4}$$

where $\|u\|_{\mathrm{TV}}$ can be considered as a L-1 norm of the first-order gradient image ∇u. The TV-based algorithm combined the steepest decent method and the projection on convex sets (POCS) to achieve the solution of (4) iteratively [6]. The steepest decent method is applied to optimize $\|u\|_{\mathrm{TV}}$ and the POCS is applied for data consistency constraints.

B. The fractional-order TV-based image reconstruction method

According to the first-order TV-based algorithm, it is straightforward that if an image is sparse with first-order gradient, and it will be also sparse with fractional-order gradient. As mentioned in the first section, the order of the regularization item is critical. Over-smoothing effect will appear when first-order term is employed. High-order term will cause speckle-like noise [9]. To achieve a good tradeoff between them, we introduce fractional-order regularization item into TV-based algorithm. The minimization problem can be written as

$$\min \|u\|_{\mathrm{FTV}} \quad \text{subject to } u \geq 0,\ Au = f, \tag{5}$$

where $\|u\|_{\mathrm{FTV}}$ can be considered as a L-1 norm of the fractional-order gradient image $\nabla^\alpha u$. The other notations in (5) is same as (4).

Fractional-order gradient can be viewed as a generalization of the integer-order gradient composed of fractional-order derivative of different direction. For computational simplicity, we use the Grüwald-Letnikov fractional-order derivative which is defined as

$$D^\alpha s(x) = \lim_{h \to 0^+} \frac{\sum_{k \geq 0} (-1)^k C_k^\alpha s(x - kh)}{h^\alpha}, \quad \alpha > 0, \tag{6}$$

where $s(x)$ is a real function, $C_k^\alpha = \Gamma(\alpha+1)/[\Gamma(k+1)\Gamma(\alpha - k + 1)]$ denotes the generalized binomial coefficient and $\Gamma(x)$ is the Gamma function. When $\alpha = 1$, for $k \geq 2$, (6) will be the first-order derivative. We choose grid length $h = 1$, so we can obtain the finite fractional-order differential operator as follow:

$$\Delta^\alpha s(x) = \sum_{k=0}^{K-1} (-1)^k C_k^\alpha s(x - k). \tag{7}$$

Especially, when $\alpha = 1$ and $K = 2$, (7) will be the first-order backward difference.

For any 2-dimensional images, the discrete fractional-order gradient $\nabla^\alpha u$ is given by

$$\nabla^\alpha u = (\Delta_x^\alpha u, \Delta_y^\alpha u), \tag{8}$$

with

$$\begin{aligned} \Delta_x^\alpha u &= \sum_{k=0}^{K-1} (-1)^k C_k^\alpha u_{x-k,y}, \\ \Delta_y^\alpha u &= \sum_{k=0}^{K-1} (-1)^k C_k^\alpha u_{x,y-k}, \\ x &= 1, 2, \ldots, M,\ y = 1, 2, \ldots, N. \end{aligned} \tag{9}$$

Setting $k = n \leq K-1$, the previous $n+2$ approximate extensive backward difference of the fractional-order differentials with respect to the negative x- and y-axis of $u_{x,y}$ can be expressed as (10) and (11).

For simplicity, we only use four directions in the fractional-order masks for the computation, corresponding to positive x- and y-axis, negative x- and y-axis. Let D_{x+}^α, D_{x-}^α, D_{y+}^α and D_{y-}^α denote the results for the four directions, see Fig.3.

The coefficients of the masks in Fig.3 are:

$$\begin{cases} C_{u_{-1}} = \frac{\alpha}{4} + \frac{\alpha^2}{8} \\ C_{u_0} = 1 - \frac{\alpha^2}{2} - \frac{\alpha^3}{8} \\ C_{u_1} = -\frac{5\alpha}{4} + \frac{5\alpha^3}{16} + \frac{\alpha^4}{16} \\ \vdots \\ C_{u_k} = \frac{1}{\Gamma(-\alpha)} \left[\frac{\Gamma(k-\alpha+1)}{(k+1)!} \cdot \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right. \\ \left. + \frac{\Gamma(k-\alpha)}{k!} \cdot \left(1 - \frac{\alpha^2}{4} \right) + \frac{\Gamma(k-\alpha-1)}{(k-1)!} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \\ \vdots \\ C_{u_{n-2}} = \frac{1}{\Gamma(-\alpha)} \left[\frac{\Gamma(n-\alpha-1)}{(n-1)!} \cdot \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right. \\ \left. + \frac{\Gamma(n-\alpha-2)}{(n-2)!} \cdot \left(1 - \frac{\alpha^2}{4} \right) + \frac{\Gamma(n-\alpha-3)}{(n-3)!} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \\ C_{u_{n-1}} = \frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(1 - \frac{\alpha^2}{4} \right) \\ + \frac{\Gamma(n-\alpha-2)}{(n-2)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \\ C_{u_n} = \frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \end{cases} \tag{12}$$

III. Experiments

To evaluate the performance of the proposed FTV method, we used some numerical experiments to compare our method with other methods, including FBP, EM and TV-POCS. The fan-beam geometry configuration of the experiments is all set as [6]. Using the Abdomen phantom shown in Fig.4(a) [16], we obtain the projection data with 20 views which are uniformly distributed over all 360 degree range. The parameters of other methods are set as recommended. In addition, the root mean squared error (RMSE) is employed to give a quantitative measurement. The reconstruction results are provided in Fig.4.

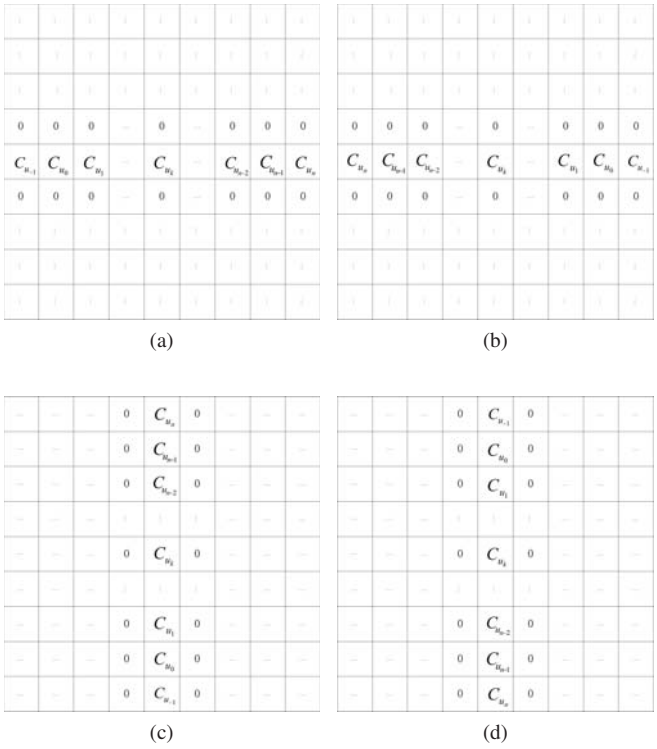
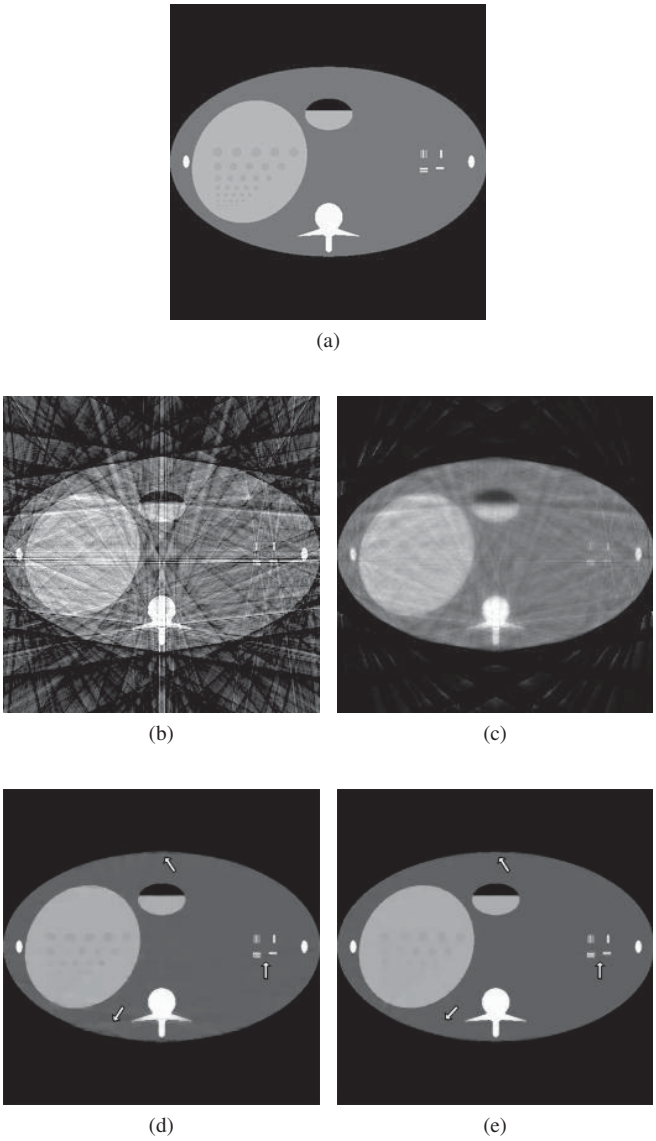


Fig. 3: Fractional-order differentials masks of four directions. (a)D_{x+}^α, (b)D_{x-}^α, (c)D_{y+}^α, (d)D_{y-}^α.

Due to incompleteness of the projection data , the classical FBP can not achieve a good visual effect and the artifacts are severe (see Fig.4(b)). Also it can be seen that after processed by other methods, the artifacts are suppressed by different degrees. Especially, in Fig.4(d) and Fig.4(e), the artifacts are

$$\begin{aligned} \Delta_x u \cong & \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) u_{x+1,y} + \left(1 - \frac{\alpha^2}{2} - \frac{\alpha^3}{8} \right) u_{x,y} \\ & + \frac{1}{\Gamma(-\alpha)} \times \sum_{k=1}^{n-2} \left[\frac{\Gamma(k-\alpha+1)}{(k+1)!} \cdot \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) + \frac{\Gamma(k-\alpha)}{k!} \cdot (1 - \frac{\alpha^2}{4}) + \frac{\Gamma(k-\alpha-1)}{(k-1)!} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \times u_{x-k,y} \\ & + \left[\frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(1 - \frac{\alpha^2}{4} \right) + \frac{\Gamma(n-\alpha-2)}{(n-2)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \times u_{x-n+1,y} + \frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) u_{x-n,y} \end{aligned} \tag{10}$$

$$\begin{aligned} \Delta_y u \cong & \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) u_{x,y+1} + \left(1 - \frac{\alpha^2}{2} - \frac{\alpha^3}{8} \right) u_{x,y} \\ & + \frac{1}{\Gamma(-\alpha)} \times \sum_{k=1}^{n-2} \left[\frac{\Gamma(k-\alpha+1)}{(k+1)!} \cdot \left(\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) + \frac{\Gamma(k-\alpha)}{k!} \cdot (1 - \frac{\alpha^2}{4}) + \frac{\Gamma(k-\alpha-1)}{(k-1)!} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \times u_{x,y-k} \\ & + \left[\frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(1 - \frac{\alpha^2}{4} \right) + \frac{\Gamma(n-\alpha-2)}{(n-2)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) \right] \times u_{x,y-n+1} + \frac{\Gamma(n-\alpha-1)}{(n-1)!\Gamma(-\alpha)} \cdot \left(-\frac{\alpha}{4} + \frac{\alpha^2}{8} \right) u_{x,y-n}. \end{aligned} \tag{11}$$



(a) (b) (c) (d) (e)

Fig. 4: The reconstruction results of Abdomen phantom. (a)is original image, (b)is reconstructed by FBP and $RMSE = 0.1479$, (c)is reconstructed by EM and $RMSE = 0.0388$, (d)is reconstructed by TV and $RMSE = 0.0132$, (d)is reconstructed by FTV ($\alpha = 1.4$) and $RMSE = 0.0099$.

almost eliminated except for some small details and the RMSE is much better than FBP and EM. Compared with first-order TV, FTV has a better ability of structure preservation between the high contrast and low contrast regions. The strength of artifacts marked by white arrows in Fig.4(e) gets slightly weaker compared with Fig.4(d). The other issue we should mention here is about the order of our method. The order in our experiments is set manually with experience. Generally speaking, α should set between 1 and 2. If $\alpha = 1$, our method will be equal to first-order TV [6] and if $\alpha = 2$, the performance will be similar with fourth-order diffusion [7].

IV. Conclusion

In this paper, we proposed a fractional-order total variation algorithm for the constrained minimization reconstruction for few-view CT problems. The proposed method use fractional-order total variation regularization term instead of first-order total variation. Since first-order TV usually smoothes structure information, we adjust the order of regularization term to fit the data fidelity better. The numerical experiments demonstrate that the proposed method is more efficient than other current methods and as a result reconstruction image converges to a better visual effect. Our future work will focus on the adaptive selection of the order and exploring the relationship between the order and the processing image.

Acknowledgment

This work is supported by the grants from the National Science Foundation of China (60972131) and the Foundation of Sichuan University Early Career Researcher Award (2012SCU11036, 2012SCU11070).

References

[1] D. J. Brenner and E. J. Hall, "Computed tomography - An increasing source of radiation exposure," *New Eng. J. Med.*, vol. 357, pp. 2277-2284, 2007.

[2] R. Gordon, R. Bender and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471-81, 1970.

[3] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrason. Imaging*, vol. 6, no. 1, pp. 81-94, 1984.

[4] A. P. Dempster, N. M. Laired and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B*, vol. 39, no. 1, pp. 1-38, 1977.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.

[6] E. Y. Sidky, C. M. Kao and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. Technol.*, vol. 14, no. 2, pp. 119-39, 2006.

[7] Y.-L. You and M. Kaveh, "Fourth-order partial differential equationsfor noise removal," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1723-1730, 2000.

[8] T. F. Chan, A. Marquina, and P. Mulet, "High-order total variation based image restoration," *SIAM J. Sci. Comput.*, vol. 22, no. 2, pp. 503-516, 2000.

[9] M. Lysaker, S. Osher, and X.-C. Tai, "Noise removal using smoothed normals and surface fitting," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1345-1357, 2004.

[10] J. Zhang and Z. Wei, "A class of fractional-order multi-scale variational models and alternating projection algorithm for image denoising," *Appl. Math. Model.*, vol 35, no. 5, pp. 2516-2528, 2011.

[11] J. Bai and X. Feng, "Fractional-order anisotropic diffusion for image denoising," *IEEE Trans. Image Process.* vol. 16, no. 10, pp. 2492-2502, 2007.

[12] Y. Zhang, Y.-F. Pu, J.-R. Hu and J.-L. Zhou, "A class of fractional-order variational image inpainting models," *Appl. Math. Inform. Sci.*, vol. 6, no. 2, pp. 229-306, 2012.

[13] Y. Zhang, Y.-F. Pu, J.-R. Hu, Y. Liu and J.-L. Zhou, "A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting," *J. X-ray Sci. Technol.*, vol. 19, no. 3, pp. 373-384, 2011.

[14] Y. Zhang, Y.-F. Pu, J.-R. Hu, Y. Liu, Q.-L. Chen and J.-L. Zhou, "Efficient CT metal artifacts reduction based on fractional-order curvature diffusion," *Comput. Math. Method Med.*, 2011, 173748.

[15] Y.-F. Pu, J.-L. Zhou, and X. Yuan, "Fractional differential mask:a fractional differential-based approach for multiscale texture enhancement," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 491-511, 2010.

[16] FORBILD Phantom [Online]. Available: http://www.imp.uni-erlangen.de/phantoms/index.htm

Model-based X-ray spectrum estimation from scanning data of CT phantoms

Huitao Zhang and Peng Zhang

***Abstract*—X-ray spectrum plays an important role in dual spectral X-ray CT imaging, CT beam hardening correction, quantitative CT analysis and so on. The conventional methods estimate the spectrum from a set of transmission data measured directly for different thicknesses of step-wedge phantoms. In this paper, we propose a novel estimation method of X-ray spectrum. The proposed method has two features. First, we restore the dependency of the transmission intensity with the thickness of attenuation materials from the CT data of simple phantoms. This not only simplifies the phantom production and the measurement process, but also can restore a more accurate dependency of the transmission intensity with the thickness of materials. Second, we suggest an improved parameter spectrum model for the spectrum estimation, in which the effect of the anode material is considered. However the spectrum estimation under such model comes down to a nonlinear estimation of multiple parameters. Hence we develop an alternative iteration algorithm to solve it. The results of numerical experiments with several simulation data suggest that the proposed method is capable to reconstruct X-ray spectra more accurate and robust for both bremsstrahlung and characteristic photons, compared to some transmission measurement methods.**

***Keywords*—X-ray spectrum estimation; parameter spectrum model; CT scanning data; alternative iteration algorithm**

I. Introduction

X-ray spectrum plays an important role in dual spectral X-ray CT imaging, CT beam hardening correction, quantitative CT analysis and so on. Due to the high photon flux produced by CT X-ray tubes, the spectrum is difficult to be directly measured. Therefore, various methods have been developed to estimate X-ray spectrum.

The conventional methods estimate the X-ray spectrum from transmission data measured directly for a set of thicknesses of step-wedge phantoms made by some materials, for instance, water (or polycarbonate) and aluminum for medical CT while iron and copper for industrial CT. After discretization, the spectrum estimation problem was converted to a linear system, and expectation maximization (EM) method was used to solve it [1, 4]. As mentioned in [4], the EM method cannot recover the details of the spectrum such as a characteristic peak if the initial guess does not contain a peak at the same energy. In order to recover X-ray spectrum with the details of both bremsstrahlung and characteristic peak, a parameter model of X-ray spectrum was suggested in [2] and the estimated characteristic peak was improved, but the bremsstrahlung photons did not match with original one well. Another method is to employ scattering measurements to reconstruct the X-ray spectrum [5]. However this method needs to use an extra detector to measure the scattered beam.

In this paper, we propose a novel estimation method of X-ray spectrum. The proposed method has two features. First, we employ the CT scanning data of some phantoms with simple structures made by water (or polycarbonate) and aluminum to estimate X-ray spectrum. Second, we suggest a new parameter model of X-ray spectrum, which provides a mechanism to exactly estimate both bremsstrahlung and characteristic peak. As mention above, in conventional spectrum estimation methods from transmission measurements, the step-wedge phantoms such as polycarbonate and aluminum were used to produce different attenuation levels. In this way, just a few to a dozen thicknesses of phantoms could be measured, which will cause the related linear system to be ill-conditioned. On the other hand, both the production of the step-wedge phantoms and their geometric placements need to be accurate and correct. However, we reconstruct a polynomial approximation of the inverse function of the transmission intensity with the thickness of attenuation materials, from the CT data of the phantoms with simple structures.

In order to estimate both bremsstrahlung and characteristic photons accurately, we suggest an improved parameter spectrum model for the spectrum estimation, in which the effect of the anode material is considered. Different from the conventional methods in [1, 4], the spectrum estimation problem with the parameter model comes down to a nonlinear estimation problem. Hence we develop an alternative iteration algorithm to solve this problem.

The results of numerical experiments with several simulation data suggest that the proposed method is capable to reconstruct X-ray spectra more accurate and robust for both bremsstrahlung and characteristic photons, compared to some transmission measurement methods.

II. Models and Methods

In this section, we introduce the two related models and methods, and then propose our model and method. The transmission measurements of polychromatic X-ray intensity can be formularized as

$$I(L) = \int S(E) e^{-\int_L \mu(\boldsymbol{x},E)\, dl} dE, \tag{1}$$

where E is the a photon energy; $\mu(\boldsymbol{x}, E)$ is linear attenuation coefficient; L represent the path length of X-ray through the object; $S(E)$ is a normalized effective spectrum related to

The authors are with The CT Laboratory, School of Mathematics, Capital Normal University, Beijing, 10048, China; email: zhanght@mail.cnu.edu.cn

X-ray emission spectrum, the scintillator of detector, and the material and thickness of filter, that is $\int S(E)dE = 1$.

When the detected object consists of single even material, the transmission measurements of polychromatic X-ray intensity can be written as

$$I(h) = \int S(E)e^{-h\mu(E)}dE, \tag{2}$$

where h represents the length of X-ray through the object.

A. Linear Model and Method

For conventional spectrum estimation from transmission measurements, $I(h)$ is directly measured for a set of a given thicknesses $(h_1, h_2, \cdots, h_M)$. Then, after discretizing $S(E)$, Eq. (2) is converted to a linear system

$$\boldsymbol{As} = \boldsymbol{I} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{s} = (S(E_1), S(E_2), \cdots, S(E_N))$ is unknown spectrum vector , $\boldsymbol{\varepsilon}$ is measuring error, $\boldsymbol{I} = (I(h_1), I(h_2), \cdots, I(h_M))$ is measurement vector, $\boldsymbol{A}$ is a known matrix with elements defined as

$$A_{ij} = e^{-h_j\mu(E_i)}\triangle E_i, \tag{4}$$

$i = 1, 2, \cdots, N; j = 1, 2, \cdots, M$.

The expectation maximization (EM) method was used to solve Eq. (3). But it cannot recover the details of the spectrum such as a characteristic peak if the initial guess does not contain a peak at the same energy [1, 4].

B. Parameter Model and Method

In order to recover X-ray spectrum with the details of both bremsstrahlung and characteristic peak, [2] suggested a parameter spectrum model

$$S(E, \boldsymbol{b}, \boldsymbol{c}, d_1) = (\Phi(E, \boldsymbol{b}) + \Psi(E, \boldsymbol{c}))e^{-d_1E^{-3}}, \tag{5}$$

where $\Phi(E, \boldsymbol{b})$ represents bremsstrahlung spectrum emitted from X-ray tube,

$$\Phi(E, \boldsymbol{b}) = b_0 + b_1E + b_2E^2 + b_3E^3, \tag{6}$$

$\boldsymbol{b} = (b_0, b_1, b_2, b_3)$; $\Psi(E, \boldsymbol{c})$ represents characteristic spectrum emitted from X-ray tube,

$$\Psi(E, \boldsymbol{c}) = \sum_{i=1}^{m}(c_i\delta(E - E_i)), \tag{7}$$

$\boldsymbol{c} = (c_1, c_2, \cdots, c_m)$, δ is Dirac function, and $e^{-d_1E^{-3}}$ represents attenuation rate when X-ray photons penetrate through the intrinsic window of X-ray tube and filters. The author estimated Tungsten Target X-ray spectrum from a few transmission measurement data in [2]. This method could improve the estimation of the characteristic peak, but could not well restore the bremsstrahlung part.

C. Our Model and Method

Now we employ CT data of simple phantoms to estimate the X-ray spectrum, instead of using transmission measurement data of the step-wedge phantoms. Let $\mu(E)$ and $f(\boldsymbol{x})$ represent the linear attenuation coefficient and density distribution of a CT phantom respectively, and $\mathfrak{R}_{u,\beta}f$ represent the fan-beam projection transformation of $f(\boldsymbol{x})$, where u is a coordinate of detector and β is the rotation angle of the CT turntable counterclockwise. Then the polychromatic CT data of the phantom is described as follows

$$I(u, \beta) = \int S(E)e^{-\mu(E)\mathfrak{R}_{u,\beta}f(\boldsymbol{x})}dE. \tag{8}$$

In order to recover X-ray spectrum with the details of both bremsstrahlung and characteristic peak more accurate, we suggest an improved parameter spectrum model as follows

$$S(E, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}) = (\Phi(E, \boldsymbol{b}) + \Psi(E, \boldsymbol{c}))\Theta(E, \boldsymbol{d}), \tag{9}$$

where

$$\Theta(E, \boldsymbol{d}) = e^{-d_1E^{-3}-d_2\mu_w(E)}, \tag{10}$$

$\boldsymbol{d} = (d_1, d_2)$, in which we add $d_2\mu_w(E)$, the effect of anode material. We will verify in the next section that the improved parameter model will match with the actual X-ray spectrum much better.

In the following, we explain how to employ the CT data of the phantom and the parameter spectrum model Eq. (9) to estimate spectrum. We convert this problem into two subproblems:

(i) To reconstruct an approximated relation function $h = H(I)$, from the CT data $I(u, \beta)$ and the known structure of the phantoms;

(ii) To estimate the X-ray spectrum form the following optimization problem

$$(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}) = \arg\ \min\{\int(I - \int S(E, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d})e^{-\mu(E)H(I)}dE)^2dI\}. \tag{11}$$

First, we consider the subproblem (i).

Set $h(p) = a_0 + a_1p + a_2p^2 + a_3p^3 + a_4p^4$, where $p = -\ln(I(u, \beta)/I_0)$. Similar to [3]., we can determine $a_n(n = 0, 1, 2, 3, 4)$ by minimizing

$$\int w(\boldsymbol{x})((f(\boldsymbol{x}) - \sum_{n=0}^{4} a_n(\mathfrak{F}p^n)(\boldsymbol{x}))^2)d\boldsymbol{x}, \tag{12}$$

where $w(\boldsymbol{x})$ is a weight image which can be used to accentuate certain image areas, $\mathfrak{F}$ is the filter back-projection operator. Then we obtain $h = H(I)$ defined as

$$H(I) = \sum_{n=0}^{4} a_n(-\ln(I))^n. \tag{13}$$

Next, we consider the subproblem (ii).

After discretizing E and I, optimization problem of Eq. (11) becomes

$$(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}) = \textbf{arg min}\|\boldsymbol{As}(\boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}) - \boldsymbol{I}\|^2, \tag{14}$$

where $\boldsymbol{s(b,c,d)} = (S(E_1,\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}), S(E_2,\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}),\cdots, S(E_N,\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}))$, $\boldsymbol{I} = (I_1, I_2,\cdots, I_M)$, and the element of $\boldsymbol{A}$ is defined as

$$A_{ij} = e^{-H(I_j)\mu(E_i)}\triangle E_i, \tag{15}$$

while $i = 1,2,\cdots,N; j = 1,2,\cdots,M$. We develop an alternative iteration algorithm to solve Eq. (14), i.e., by alternatively using the iteration formula of solving

$$\check{\boldsymbol{s}} = \textbf{arg min}\|\boldsymbol{A}\check{\boldsymbol{s}} - \boldsymbol{I}\|^2 \tag{16}$$

and

$$\boldsymbol{(b,c,d)} = \textbf{arg min}\|\boldsymbol{s(b,c,d)} - \check{\boldsymbol{s}}\|^2. \tag{17}$$

We summary the steps of the algorithm as follows.

1 Selecting an initial spectrum vector $\boldsymbol{s^{(0)}}$

2 Suppose $\boldsymbol{s^{(n)}}$ is known after nth interations, we determine $\check{\boldsymbol{s}}^{\boldsymbol{(n)}}$ by the iteration formula of SART method for solving the problem $\boldsymbol{A'A\check{s}^{(n)} = A'I}$ (remark: $\check{\boldsymbol{s}}^{\boldsymbol{(n)}}$ is a solution of Eq. (16) if and only if $\boldsymbol{A'A\check{s}^{(n)} = A'I}$).

3 Determining $\boldsymbol{s(b,c,d)^{(n)}}$ by the iteration formula of Levenberg-Marquardt method for solving the nonlinear optimization problem

$$\textbf{min}\|\boldsymbol{s(b,c,d)^{(n)} - \check{s}^{(n)}}\|^2 \tag{18}$$

4 Calculating $\boldsymbol{s^{(n+1)}} = \check{\boldsymbol{s}}^{\boldsymbol{(n)}} + \alpha(\boldsymbol{s(b,c,d)^{(n)}} - \check{\boldsymbol{s}}^{\boldsymbol{(n)}}), \alpha \in (0,1)$

5 Returning step 2, until $\|\boldsymbol{As^{(n+1)} - I}\|$ is less than the given threshold or n reaches the liminal number.

III. Experiments

To verify our method, we use simulation data to carry out the experiments. The simulation X-ray spectrum used in experiments is obtained from the open source software Spectrum GUI [6]. The attenuation coefficient for each material is obtained from the NIST(National Institute of Standard Technology Web)[7].

A. Reconstruction of the relation $h = H(I)$

We use the simulation polychromatic CT data of the phantoms to verify our algorithm that reconstructs the relation $h = H(I)$. In experiments, we choose a pyramid-shaped aluminum phantom. The cross section to be scanned is a rectangle with a ratio of length to width 3. The choice of the length depends on the voltage of X-ray tube. Under 140 kVp, we choose the length as 3 cm.

We first reconstruct each $\Im p^n(\boldsymbol{x})$ from polychromatic CT data of the phantom, and then define the weight function $w(\boldsymbol{x})$ to be 0 at the two-pixel neighbourhood of the rectangle boundary, and to be 1 in the other part. By minimizing Eq.(12), we obtain $a_0 = 0.104578$, $a_1 = 22.004794$, $a_2 = 0.801286$, $a_3 = -0.033721$, $a_4 = 0.000609$. According to definition Eq.(13), we get $H(I) = a_0(-\ln(I)) + a_1(-\ln(I)) + a_2(-\ln(I))^2 + a_3(-\ln(I))^3 + a_4(-\ln(I))^4$, as shown in Fig.1. We can find $I = H(I)$ well match with the relationship of the polychromatic transmission measurement with the thickness of the phantom.

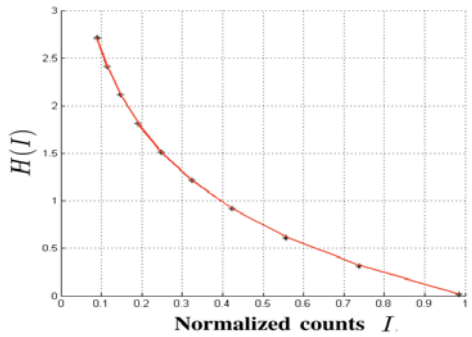


Fig. 1. The star points in the figure are the measurements of $(I(h), h)$ at the sampled thicknesses of the step-wedge phantom. The curve is the plot of the function $h = H(I)$ reconstructed by the CT data of the CT phantom with rectangle-shape.

B. Rationality of the parameter spectrum model

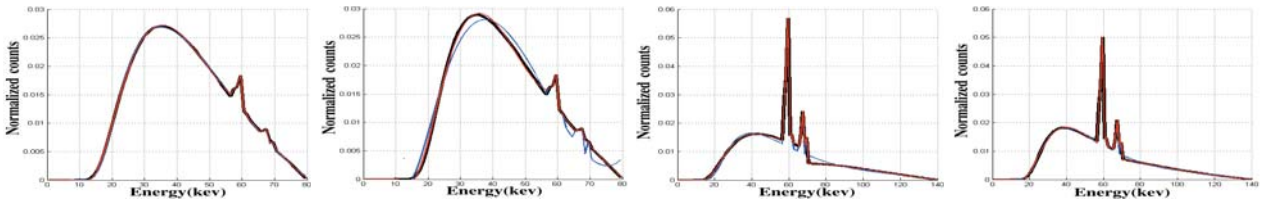
In this subsection, we demonstrate the rationality of our parameter spectrum model. We use our parameter spectrum model and the parameter spectrum model in [2] to fit some typical spectra of GE Maxiray 125 and Dunlee PX1557 under the voltages 80 and 140 kVp of X-ray tube with tungsten target. Fig.2 show the fitting results, from which one can see that the spectra fitted by our parameter spectrum model match with original ones much well, while the spectra fitted by the parameter spectrum model [2] deviate from original ones, especially at the neighbourhood of the absorption edge of the anode material.

C. Results of spectrum estimation

We use the spectra of the X-ray tube of Dunlee PX1557 at 140 kVp and 80 kVp to verify our spectrum estimation method, and compare the estimated results with those of the EM method and the method in [2]. The cross section of the CT phantom we use here is the same as that in the subsection A. The CT data are simulated with 1730 bins for each view and 720 views for full turn. The comparisons of original spectra with the spectra estimated show in Fig.3. The black curves are the original spectra. The voltage of X-ray tube for (a)(b)(c) is 140 kVp while for (d)(e)(f) is 80 kVp respectively. The spectra (the red curves) in (a) and (d) are estimated by our method, in (b) and (e) by the method of [2], and in (c) and (f) by the EM method. All spectra estimated are the results after 500 iterations. One can find that the spectra estimated by our method match the original ones better than those estimated by other two methods. All data here are noise free. In fact, we have made the experiments using the CT data with Poisson noise. The results show that our spectrum estimation method is robust to the CT data with Poisson noise.

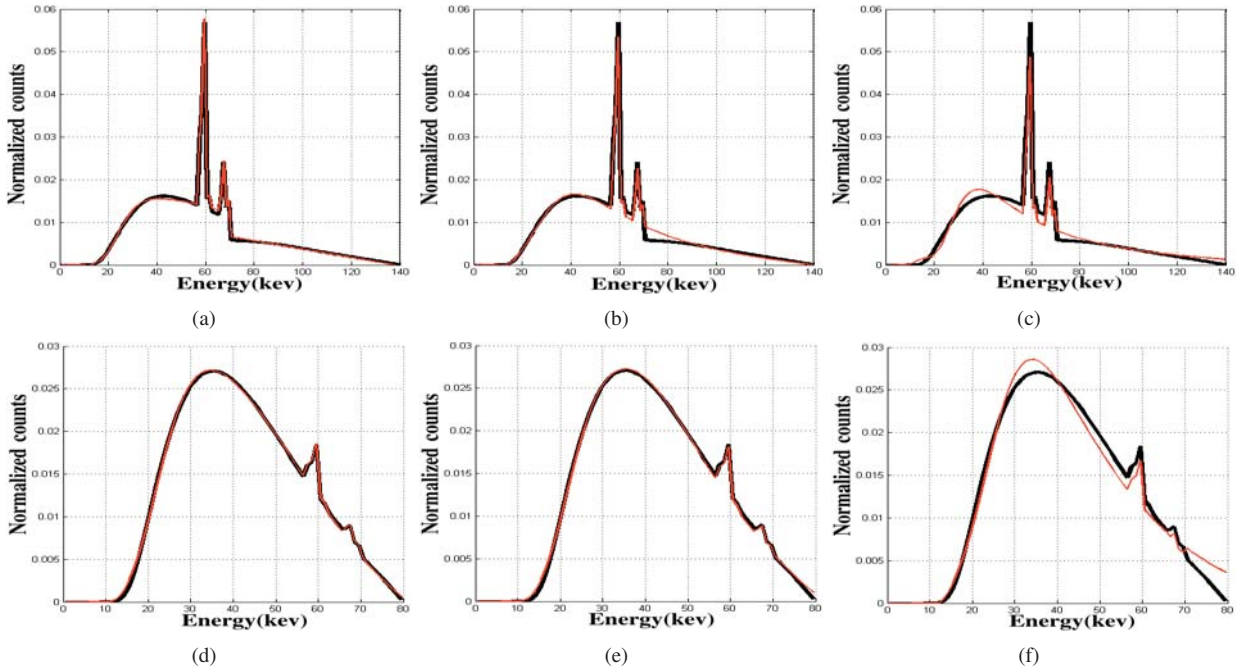
IV. Conclusion

We have proposed a novel model and the related method to estimate the X-ray spectra. Compared with transmission measurement methods of the step-wedge phantoms, our method has two benefits: (i) the CT phantoms are easy to be produced; (ii) the CT phantom need not be placed on CT turntable exactly. We have verified that our parameter spectrum model is more



(a) (b) (c) (d)

Fig. 2. The comparisons of the spectra of the different X-ray tubes with the spectra fitted by two parameter spectrum models. The black curves in (a) and (c) are the spectra of the X-ray tube of the Dunlee PX1557 at 80 kVp and 140 kVp respectively, and in (b) and (e) are the spectra of the X-ray tube of GE Maxiray 125 at 80 kVp and 140 kVp respectively. The red curves are fitted by our parameter spectrum model, and blue curves are fitted by the parameter spectrum model of [2].



(a) (b) (c)

(d) (e) (f)

Fig. 3. The comparisons of original spectra (the black curves) with the spectra estimated (the red curves) by our method in (a) and (d), by the method of [2] in (b) and (e), and by the EM method in (c) and (f). The voltage of X-ray tube for (a)(b)(c) is 140 kVp while for (d)(e)(f) is 140 kVp.

rational to describe spectra for various X-ray tubes, which is a key that our method is able to reconstruct X-ray spectra more accurate for both bremsstrahlung and characteristic photons. The results of the simulation experiments have verified that the proposed method is efficient to estimate X-ray spectra under different conditions. The further work is to check the method in real CT systems. The optimization of phantoms in material, structure and length also should be studied.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under the grants 61127003, Beijing Education Committee under the grants PHR20110509 and KZ201110028034, Beijing Natural Science Foundation under the grants 3112006.

REFERENCES

[1] X.H. Duan, J. Wang, L.F. Yu, S.A. Leng, and C.H. McCollough, CT scanner x-ray spectrum estimation from transmission measurements, Med Phys, vol. 38, (no. 2), pp. 993-997, 2011.

[2] Y. Yang, X.Q. Mou, H.J. Yu, X. Chen, Y.B. Zhang, and S.J. Tang, Model-Based Tungsten Anode X-ray Tube Spectra Reconstruction Method, Acta Electronica Sinica, vol. 10, (no. 10), pp. 015, 2010.

[3] M. Kachelrie Ss, K. Sourbelle and W.A. Kalender, Empirical cupping correction: a first-order raw data precorrection for cone-beam computed tomography, Med Phys, vol. 33, pp. 1269, 2006.

[4] E.Y. Sidky, L. Yu, X. Pan, Y. Zou, and M. Vannier, A robust method of x-ray source spectrum estimation from transmission measurements: Demonstrated on computer simulated, scatter-free transmission data, J Appl Phys, vol. 97, (no. 12470112), 2005.

[5] J.E. Fernandez, V. Scot, J. Bare, F. Tondeur, S. Gallardo, J. Rodenas, and P.L. Rossi, Reconstruction of the X-ray tube spectrum from a scattering measurement, Appl Radiat Isotopes, vol. 70, (no. 7SI), pp. 1238-1242, 2012.

[6] http://sourceforge.net/projects/spectrumgui/

[7] http://physics.nist.gov/PhysRefData/XrayMassCoef/ cover.html.

A Comparative Study of Neighborhood Filters for Artifact Reduction in Iterative Low-Dose CT

Wei Xu, Sungsoo Ha, Ziyi Zheng and Klaus Mueller

***Abstract*— Iterative CT algorithms have become increasingly popular in recent years. They have been found useful when the projections are limited in number, irregularly spaced, or noisy, which are often encountered in low-dose CT imaging. One way to cope with the associated streak and noise artifacts is to interleave a regularization objective into the iterative reconstruction framework. In this paper we investigate a number of non-linear neighborhood filters within an iterative CT framework, OS-SIRT, and compare them with total variation minimization (TVM). We find that the Non-Local Means (NLM) filter provides the best performance, in particular its patch-based variant. Further, we also compare a scheme that exploits an artifact-free reference image for even better regularization performance. Finally, we also compare the studied filters in terms of their computational efficiency with acceleration on modern GPUs.**

I. INTRODUCTION

Low dose CT imaging has been gaining considerable momentum in recent years. However, low-dose CT leads to noisy and sparse X-ray projections, which subsequently lead to significant noise and streak artifacts in the reconstructions. In these adverse conditions iterative reconstruction algorithms are more favorably applied, especially when combined with regularization. Here, the method of Total Variation Minimization (TVM), has become rather popular and has been used in many frameworks, such as ASD-POCS [5]. However, TVM is an iterative global optimization algorithm and can be costly in compute, lessen practicality in clinical practice.

We study if non-iterative filters that only operate in a local neighborhood can lead to improved results. An advantage here is that they also lend themselves very well to GPU acceleration. We specifically study and compare the bilateral filter (BLF) [6] and the non-local means filter (NLM) [1]. While the use of local neighborhood filters within an iterative CT reconstruction framework is not conceptually new, this paper's contribution is (1) a comparison of these both in terms of quality and speed, and (2) their extension into an adaptive form [3] and one that uses a prior image of the patient [11] .

II. OVERVIEW

Our reconstruction framework is fully iterative using our OS-SIRT pipeline [7][9] for reconstruction, interleaving regularization within each iteration. The regularization enforces constraints in the object domain, such as local smoothness and coherent edges, while the reconstruction ensures fidelity with the acquired data. This type of pipeline has also been used by others, but with different reconstruction algorithms and regularization schemes. All operations are accelerated on the GPU.

A. Regularization as a denoising task

In the context of mitigating artifacts in CT reconstruction, regularization is similar to the process of *denoising* in image processing. In fact, the notion of noise is quite general and can include for example streak artifacts. We may distinguish between two families of denoising strategies: (i) global optimization and (ii) local filtering. Both can be iterative, where the former seeks to improve some global objective function and the latter repeats the filtering, possibly guided by some error criterion and varying parameters along the way. In fact, the two families can be unified into a mathematical framework which gives them a common theoretical underpinning. Elad [2] shows that both derive from a solid theory of statistical estimators and regularization. More specifically, the bilateral filter emerges from the Bayesian approach as a single iteration of the Jacoby normalized diagonal steepest descent algorithm.

For the remainder, we shall adhere to the terminology of image processing where the goal is to reduce artifacts in images.

B. Regularization by local neighborhood filtering

Local neighborhood filters have become popular in image processing since they can achieve better computational performance and also afford local control. They are non-iterative (although they can be applied repeatedly) and are based on pixel-wise operations over a small neighborhood.

In contrast to global optimization such as TVM, for nonlinear neighborhood filters (NNF) the updated value at a pixel x is determined by a weighted sum of a functional mapping α of its local neighborhood W_x. This typically non-linear function α takes into account both spatial and value discrepancies with respect to x, as expressed in the following equation:

$$NNF(x, f, W_x, \alpha) = \frac{\sum_{t \in W_x} \alpha(x,t, f(x)) f(x+t)}{\sum_{t \in W_x} \alpha(x,t, f(x))} \quad (1)$$

The normalization forces the sum of pixel weights to 1. The window area W_x defining the local neighborhood can vary in size for pixels at different positions. To compute the weights of the neighborhood, a distance metric measures the similarity between the pixel at $x+t$ and the central pixel at x. Next, we use this general notation to express all filters we have studied.

The bilateral filter (BLF) [6]: The filter only considers a fixed sized neighborhood around the target pixel x, and the weighting function α_{BLF} is the product of spatial distance weight c_d and range distance weight s_r:

$$\begin{aligned} \alpha_{BLF}(x,t,f) &= c_d(t) s_r(f(x), f(x+t)) \\ c_d(t) &= G_{\sigma_d}(\|t\|) \\ s_r(f(x), f(x+t)) &= G_{\sigma_r}(|f(x) - f(x+t)|) \end{aligned} \quad (2)$$

where $G_\sigma(x)$ is the Gaussian kernel

$$G_{\sigma_r}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2}) \quad (3)$$

and σ_d and σ_r control the amount of smoothing. The function c_d acts as a domain filter to ensure spatial closeness to x such that far away pixels have no effects. On the other hand, the function s_r acts as a range filter to ensure value closeness to $f(x)$ such that the values of pixels from different nearby materials cannot

Wei Xu and Klaus Mueller are with the Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University, Stony Brook, NY 11777 USA (phone: 631-632-1524; e-mail: {wxu, mueller}@cs.sunysb.edu).

Partial support came from NSF grants 1050477, 0959979 and 1117132.

diffuse into the material represented by *x*. Similar to anisotropic diffusion it ensures that sharp edges are well preserved.

The non-local means filter (NLM) [1]: Based on the assumption that there is a high degree of redundancy in a given image, the NLM filer consults similar pixel neighborhoods (called *patches*) in disjoint image regions and average their contributions for a more stable outcome:

$$f_{NLM}(x) = NNF(x, f, W, \alpha_{NLM}) \tag{4}$$

As such, the variable *t* parameterizes the offset within the search window as before. In order to gauge the similarity of a neighborhood patch at *x*+*t* with the neighborhood at *x*, the corresponding pixel-differences are weighted by a Gaussian kernel G_a with standard deviation σ_a, inside the patch area *P*:

$$\alpha_{NLM}(x,t,f) = \exp\left(-\frac{\sum_{p\in P} G_{\sigma_a}(t)\left|f(x+p)-f(x+t+p)\right|^2}{h^2}\right) \tag{5}$$

h acts as a filtering parameter which, when increased, allows for more dissimilar patches to contribute to the smoothing.

The adaptive NLM filter (ANLM) [3]: In the NLM filter, the search window has typically a constant, pre-set size throughout the image. However, picking a good size of the NLM search window can be challenging, especially when noise levels and patterns are not spatially invariant, which is most often the case. Hence it is more appropriate to locally adapt the window size. Kervrann and Boulanger [3] describe an iterative approach (with usually less than 4 iterations) that adaptively grows the local search window to incorporate neighborhood statistics at an increasing level of scale. The expansion is terminated once the deviation bias of the weighted smoothing grows too large (i.e. the local estimates diverge at increasing scale). We call this approach *adaptive NLM (ANLM)* since it also determines the weight of a neighborhood pixel via its patch similarity (the patch size itself is fixed) – however, the similarity measure changes as the iterations proceed. At each iterative step *i*, the smoothed image $f_{ANML(i)}(x)$ and the variance $\sigma_i^2(x)$ at position *x* of a neighborhood are calculated using the adaptive weights w_i as:

$$\begin{aligned} f_{ANLM(i)}(x) &= \sum_{t\in W_{x,i}} w_i(x,t,f_{ANLM(i-1)})\, f(x+t) \\ \sigma_i^2(x) &= \sigma_0^2 \sum_{t\in W_{x,i}} w_i(x,t,\mu_{i-1})^2 \end{aligned} \tag{6}$$

Here, $W_{x,i}$ is the current neighborhood size and σ_0 is the initial standard deviation, estimated from the input image (more detail is provided in [3]). The current $f_{ANML(i)}(x)$ and $\sigma_i^2(x)$ then serve as input to compute the weights for the next iteration: (7)

$$w_i(x,t,\mu_i) = \frac{\exp(-dist(f_{ANLM(i-1)}(x), f_{ANLM(i-1)}(x+t))/h^2)}{\sum_{t\in W_{x,i}} \exp(-dist(f_{ANLM(i-1)}(x), f_{ANLM(i-1)}(x+t))/h^2)}$$

$$dist = \sum_{p\in P}\left(f_{ANLM(i-1)}(x+p) - f_{ANLM(i-1)}(x+t+p)\right)^2\left(\frac{1}{2\sigma_{i-1}^2(x+p)} + \frac{1}{2\sigma_{i-1}^2(x+t+p)}\right)$$

There are five parameters: the initial noise variance $\sigma_0^2(x)$, the patch size *P*, the parameter *h* (and a factor ρ), and the maximal number of iterations *N*. The $\sigma_0^2(x)$ can be automatically generated through robust estimation in the image. For *P*, we found a size of 7×7 practical in most cases. The other parameters were relatively insensitive to change within a normal range.

The reference-based NLM filter (RNLM) [11]: Often prior scans of the patient are available which could be used as an external site for patch-based neighborhood matching. This gives rise to the following equations:

$$\begin{aligned} f_{RNLM}(x) &= NNF_{RNLM}(x, f, f_r, f_{rn}, W, \alpha_{RNLM}) \\ NNF_{RNLM}(x, f, f_r, f_{rn}, W_x, \alpha) &= \frac{\sum_{t\in W_x} \alpha(x,t,f,f_{rn}(x))\, f_r(x+t)}{\sum_{t\in W_x} \alpha(x,t,f,f_{rn}(x))} \\ \alpha_{RNLM}(x,t,f,f_{rn}) &= \exp\left(-\frac{\sum_{p\in P} G_{\sigma_a}(t)\,|\,f(x+p)-f_{rn}(x+t+p)|^2}{h^2}\right) \end{aligned} \tag{8}$$

Here, f_r is the reference image containing similar features as the image *f* to be denoised, f_{rn} is the same reference image now augmented with similar artifact statistics, and *t* is some offset to *x* that locates areas with similar features. The value of *t* could be determined by a rough registration or an approximate feature matching using block-based histograms, etc. Our work in [11] used this filter to restore an image reconstructed with filtered backprojection. In the research presented here we use it as a regularization operator in an iterative reconstruction scheme.

III. RESULTS

We interfaced an NVIDIA GTX 480 GPU with an Intel 2 Quad CPU @ 2.66GHz host processor. For testing, we employed the NIH Visible Human's torso (size 256^3) which has a prominent spine structure with different bone sizes and small structures and a brain dataset (NIH Visible Human brain, size 256^3) which also has some finer structures. We used a high-quality X-ray simulator to obtain various projection sets for torso and brain.

In our experiments, we explore the various regularization schemes within the interleaved reconstruction pipeline, for both the few-view and the noisy projection scenarios. We found the best parameters for each filter via experimentation.

A. Qualitative and quantitative comparison: torso dataset

We simulated 180 uniformly distributed projections over a half-circle trajectory. For the few-view case we selected every 9th projection from the set, yielding a total of 20 projections. Then, for each of the 4 regularization schemes (BLF, TVM, NLM, and ANLM), we interleaved regularization with OS-SIRT (10 subsets) and ran this pipeline for a total of 200 iterations. For the second series of experiments, we added significant Gaussian noise (SNR=10) to all 180 projections and ran the same pipeline again, but this time for only 20 iterations since this yields about the same number of updates as the few-view case (this much noise typically also causes the reconstruction procedure to diverge when the noisy projections are not pre-filtered).

The results of these two experiments are shown in Figure 1 along with the corresponding parameter settings and the best E-CC metric scores they could achieve. The E-CC is a perceptual quality metric and was introduced in [10] – it measures the cross-correlation (CC) of an edge-filtered image. We provide ROI zoomed results for two critical regions, spine and lung. The left-most full-body reconstructions were obtained without regularization. The first observation we make is that streaks seem to be easier to remove than heavy noise – the E-CC obtained with regularization is roughly 12-15% higher for the former for all regularization schemes. We also readily observe that all filters can reduce streaks and noise, recovering some structural parts which can be hardly seen in the non-filtered result. In the following we focus our detailed discussion on the spine – similar observations can also be made for the lung.

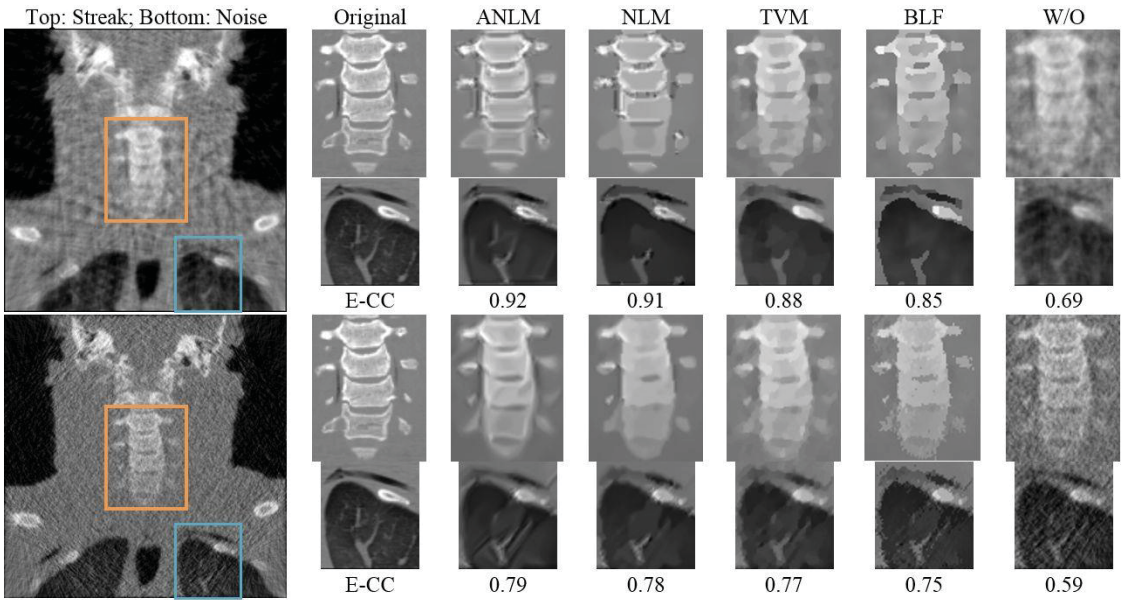


Figure 1. Torso dataset, reconstructed with the interleaved regularization pipeline both for the few-view (top, 20 projections) and noisy (bottom, 180 projections, SNR 10) scenarios. Zoomed results for two critical regions are shown, indicated by the orange (spine) and blue (lung) boxes in the left-most reconstructions obtained without regularization. The reconstructions appear ordered according to their E-CC scores.

The BLF and TVM perform quite similarly, but while the BLF keeps sharper edges and provides better streak removal than TVM, it also gives the image a more binary look signified by abrupt changes along adjacent varying-intensity areas. TVM, on the other hand, has smoother transitions here, and it also seems to perform better with noise. However, neither of the filters is able to recover more subtle features.

The two NLM-based methods successfully master the problems encountered with BLF and TVM – both recover the gaps separating the individual vertebrae. The shape and structure of the vertebrae is also better described, delineating the bony shell around the vertebrae body well. However, for the noisy projections case, the ANLM filter is the only one to do so.

B. Qualitative comparison: brain dataset

We used the same conditions as for the torso dataset (20 projections for the few-view case, 180 projections with SNR 10 Gaussian noise added for the noisy case) and the same regularized construction strategy (200 iterations with interleaved OS-SIRT 10 for the few-view case, 20 iterations of OS-SIRT 10 for the noisy projection case).

Figure 2a shows the results we obtained for the few-view case. We observe that in terms of sharpness and detail preservation ANLM and NLM have similar outcomes, but that the ANLM better preserves the small structures pointed by the arrow in the Original image. We further observe that the BLF produces slightly sharper and detailed images than TVM, but not quite as good as the NLM filter. The figure also examines the result obtained with the RNLM filter. Here we explored two different strategies (i) apply the reference image-based regularization only once (after the final iteration step), and (ii) apply it in an interleaved fashion. It can be clearly observed that the interleaved RNLM scheme preserves detail much better and restores some fine detail that the ANLM filter cannot, especially some of the interior detail of the bone structures. This fine detail is just not expressed at a strength that is sufficient enough for the (A) NLM filter to restore it from the patches found in the local image, making it necessary to use a clean source for these.

Figure 2b shows the results for the noise case. Here the differences of NLM and ANLM are not as profound as for the streak case. However, similar qualitative differences can be observed for the BLF and TVM, as well as for the two RNLM strategies (see above). Interesting for the latter is the dark feature pointed to by the arrow in the Original image. This feature does not exist in either reference image and is instead restored using local NLM since the reference-image based matching did not return a sufficiently high sum of weights (while similar is also true for the streak case above but there the greater number of iterations also enabled a better OS-SIRT data-driven reconstruction). Nevertheless, this is a clear indicator that the RNLM scheme is very sensitive to the richness of the underlying prior and ongoing research seeks to improve on this.

C. Time performance

Table 1 lists the run times to filter images of 3 different sizes (256^2, 512^2, and 1024^2) on the GPU. We found that performing filtering in 3D did not yield any improvements so restricting our experiments to 2D is well justified. In the table we list both the timings for the non-optimized (NOPT) and the optimized (OPT) GPU implementation reported in [12]. This optimization achieve a speedup of about 1.2 for the BLF, about 4 for the NLM filter, and about 3.2 for the ANLM filter. Please note that these speedups are in addition to the two orders of magnitude speedup over a corresponding CPU implementation, as reported in [8].

To estimate the TVM performance on the GPU we used TVM GPU implementation of Pock et al. [4] as a reference. They used a NVIDIA 8800 GTX for their experiments and we report their timings in Table 2 as well. In order to make these timings comparable to ours we extrapolated them to the GTX 480 using commonly reported speedup numbers. We may add, however, that once the parameter λ grows larger, which is needed for the rather noisy data we have used here, the computation time tends to increase significantly over those listed here.

Overall there is about an order of magnitude difference in the run times for each of the filters: BLF, NLM, and ANLM, with BLF being the fastest. The TVM requires about the same time as NLM. The timing of the RNLM filter is comparable to that of the NLM filter since the matching process is similar.

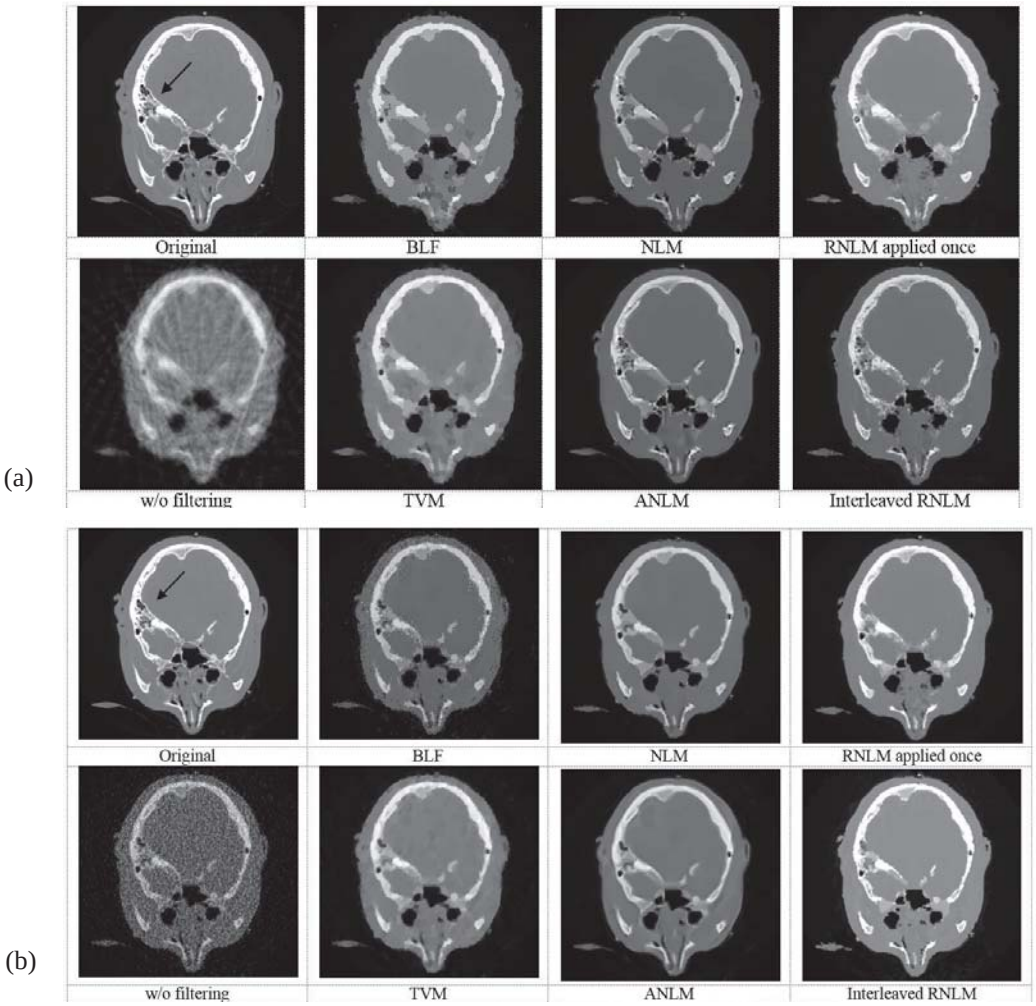


Figure 2. Brain dataset, reconstructed with the interleaved regularization pipeline for the (a) few-view scenario (20 uniformly distributed projections, 200 iterations with OS-SIRT 10). (b) noisy data (SNR 10, 180 equi-angular projections, 10 iterations with OS-SIRT 10).

IV. Conclusions

We have explored the use of local nonlinear neighborhood filtering as a non-iterative alternative to the popular TVM method for regularized CT reconstruction. Our results indicate that these types of filters can be advantageous to TVM, meeting and exceeding its capabilities.

References:

[1] A. Buades, B. Coll and J. Morel, "A Review of Image Denoising Algorithms with A New One," *Multi-scale Modeling Simulation* 4(2) 490-530 (2005).

[2] M. Elad, "On the Bilateral Filter and Ways to Improve It," *IEEE Trans. On Image Processing* 11(10) 1141–1151 (2002).

[3] C. Kervrann and J. Boulanger, "Optimal Spatial Adaptation for Patch-based Image Denoising," *IEEE Trans. on Image Processing* 15(10) 2866-2878 (2006).

[4] T. Pock, M. Unger, D. Cremers and H. Bischof, "Fast and Exact Solution of Total Variation Models on the GPU," *IEEE Computer Vision and Pattern Recognition Workshop*s 1–8 (2008).

[5] E. Sidky, X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Bio.* 53(17):4777-4807 (2008).

[6] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. Intern. Conf. on Computer Vision (ICCV)* 839-846 (1998).

[7] F. Xu, W. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard and K. Mueller, "On the Efficiency of Iterative Ordered Subset Reconstruction Algorithms for Acceleration on GPUs," *Computer Methods and Programs in Biomedicine* 98(3) 261-270 (2010).

[8] W. Xu and K. Mueller, "A Performance-Driven Study of Regularization Methods for GPU-Accelerated Iterative CT," *HPIR*, 20-23 (2009).

[9] W. Xu, F. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard and K. Mueller, "High-Performance Iterative Electron Tomography Reconstruction with Long-Object Compensation using Graphics Processing Units (GPUs)," *J. Structural Biology*, 171(2) 142-153, (2010).

[10] W. Xu and K. Mueller, "Learning Effective Parameter Settings for Iterative CT Reconstruction Algorithms," *Proc. International Meeting on Fully 3D Image Reconstruction*, 251-254 (2009).

[11] W. Xu, K. Mueller, "Efficient Low-Dose CT Artifact Mitigation Using an Artifact-Matched Prior Scan," *Medical Physics*, 39:4748-4760, 2012.

[12] Z. Zheng, W. Xu, and K. Mueller, "Performance Tuning for CUDA-Accelerated Neighborhood Denoising Filters," *HPIR* (2011).

Test Size	BLF			NLM			ANLM			TVM (from [4])	
	NOPT	OPT	Ratio	NOPT	OPT	Ratio	NOPT	OPT	Ratio	8800 GTX	GTX 480
256^2	0.65	**0.53**	1.23	51.09	**12.70**	4.02	142.32	**43.57**	3.27	17.50	6.74
512^2	2.15	**1.76**	1.22	182.49	**42.06**	4.34	374.8	**117.24**	3.20	59.60	22.95
1024^2	8.08	**6.54**	1.24	699.23	**161.25**	4.34	2072.67	**597.91**	3.47	504.10	194.15

Table 1. Wall clock time (in *ms*) of the GPU-accelerated BLF, NLM, and ANLM filters both optimized (OPT) and non-optimized (NOPT) for different image sizes. Ratio is the speedup NOPT/OPT. For TVM, the parameter λ grows larger for noisy data which further increases time.

Nonconvex optimization for improved exploitation of gradient sparsity in CT image reconstruction

Emil Y. Sidky[1], Rick Chartrand[2], Jakob S. Jørgensen[3], and Xiaochuan Pan[1]

***Abstract*—A nonconvex optimization algorithm is developed, which exploits gradient magnitude image (GMI) sparsity for reduction in the projection view angle sampling rate. The algorithm shows greater potential for exploiting GMI sparsity than can be obtained by convex total variation (TV) based optimization. The nonconvex algorithm is demonstrated in simulation with ideal, noiseless data for a 2D fan-beam computed tomography (CT) configuration, and with noisy data for a 3D circular cone-beam CT configuration.**

I. Introduction

Much recent work in iterative image reconstruction in computed tomography (CT) has focused on some form of total variation (TV) minimization, and one of the motivations for employing TV minimization is exploiting sparsity in the gradient magnitude image (GMI) to reduce sampling requirements for the CT system. TV-minimization has been demonstrated, in simulations and with real scanner data, to be effective at allowing for projection view sampling reduction. There is, however, potential to take the sparsity-exploiting principle further, because TV-minimization is an ℓ_1-based convex relaxation of an ideal, nonconvex, sparsity-exploiting optimization based on the ℓ_0-norm. To approach more closely the ℓ_0-based minimization, we develop a GMI sparsity-exploiting algorithm for CT based on an ℓ_p-norm where $p \in (0,1)$. Section II summarizes the theory and algorithm, and Sec. III shows results based on 2D and 3D CT simulations.

II. Constrained, nonconvex optimization by reweighting

We briefly state the rationale and methods for GMI exploiting CT image reconstruction with nonconvex optimization. We write the CT data model generically as a linear system

$$\mathbf{g} = X\mathbf{f}, \tag{1}$$

where $\mathbf{f}$ is the image vector comprised of voxel coefficients, X is the system matrix generated by some approximation to projection of the voxels, and $\mathbf{g}$ is the data vector containing the estimated projection samples. The model can be applied equally to 2D and 3D geometries, and we note that there are many specific forms to this linear system depending on sampling, image expansion elements, and approximation of continuous fan- or cone-beam projection.

[1]The University of Chicago, Department of Radiology MC-2026, 5841 S. Maryland Avenue, Chicago, IL 60637. Corresponding author: Emil Y. Sidky, E-mail: sidky@uchicago.edu. [2]Theoretical Division, T-5, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545. [3]Technical University of Denmark, Department of Applied Mathematics and Computer Science, Matematiktorvet, bygning 303B, 2800 Kgs. Lyngby, Denmark.

For the present work, we focus on CT configurations with sparse angular sampling, where the sampling rate is too low for Eq. (1) to have a unique solution. In this situation, there has been much interest in exploiting GMI sparsity of the object to narrow the solution space of Eq. (1) and potentially obtain an accurate reconstruction from under-sampled data. The formulation of this idea results in a nonconvex constrained optimization:

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_0 \quad \text{such that } \mathbf{g}_{\text{data}} = X\mathbf{f}, \tag{2}$$

where the argument of the ℓ_0-norm is the voxel-wise magnitude of the image spatial gradient; the linear operators ∂_x, ∂_y, and ∂_z are matrices representing finite differencing in their respective labeled directions; the numerical gradient of the image is formed by, $\nabla \mathbf{f} = [\partial_x \mathbf{f}, \partial_y \mathbf{f}, \partial_z \mathbf{f}]^T$ (2D is obtained by deleting the third component); the ℓ_0-norm counts the number of non-zero components in the argument vector; and $\mathbf{g}_{\text{data}}$ is the available projection data. In words, this optimization seeks the image $\mathbf{f}$ with the lowest GMI sparsity while agreeing exactly with the data.

The optimization problem in Eq. (2) does not lead directly to a practical image reconstruction algorithm, because, as of yet, no large scale solver is available for this problem. Also, the equality constraint, requiring perfect agreement between the available and estimated data, makes no allowance for data inconsistency. In working toward developing a practical image reconstruction algorithm different relaxations of Eq. (2) have been considered. One such relaxation is

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_p^p \quad \text{such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \leq \epsilon, \tag{3}$$

where the ℓ_0-norm is replaced by the ℓ_p-norm,

$$\|\mathbf{v}\|_p^p \equiv \sum_i |v_i|^p,$$

and the data equality constraint is relaxed to an inequality constraint with data-error tolerance parameter ϵ. An important strategy, which has been studied extensively in Compressive Sensing [1], is to set $p = 1$, which corresponds to TV-minimization. This, on the one hand, maintains some of the sparsity seeking features of Eq. (2) and, on the other hand, leads to a convex problem, which has convenient features for algorithm development. For example, a local minimizer is a global minimizer in convex optimization.

Another interesting option for GMI sparsity-exploiting image reconstruction is to consider Eq. (3) for $0 < p < 1$. Such a choice for p leads to nonconvex optimization, which can allow for greater sampling reduction than the $p = 1$ case while maintaining highly accurate image reconstruction. These gains intuitively stem from the fact $p < 1$ is closer to the ideal sparsity-exploiting case of $p = 0$; the catch, however, is on the algorithmic side where one has to deal with potential local minima, which are not part of the global solution set. Despite this potential difficulty, practical algorithms based on this nonconvex principle are available [2,3], and gains in sampling reduction for various imaging systems have been reported for both simulated and real data cases. For X-ray tomography, use of this nonconvex strategy has shown promising results [4,5], but the algorithms proposed in those works for CT are only motivated by the optimization problem in Eq. (3) and are not accurate solvers of this problem. An accurate solver is important for theoretical studies of CT image reconstruction with under-sampled data and may also aid in developing algorithms for limited-data tomographic devices.

For CT, one of the barriers to developing an efficient and accurate solver for Eq. (3) in the nonconvex $p < 1$ case, is that it is already challenging to develop such a solver for the convex $p = 1$ case. In order to handle the latter convex, but non-smooth case, we have been interested in an alternate line of optimization problems, where the salient image metrics are written as constraints instead of in an objective function. It is a strategy similar to the set theoretic approach presented in Ref. [6]; the algebraic reconstruction technique (ART) is a specific realization of this strategy; and this type of approach can be useful for nonconvex constraint sets [7]. For the alternate, constraint-based optimization problem there are efficient, large-scale solvers recently available [8,9].

Returning to GMI sparsity-exploiting image reconstruction, we employ an approach developed in Ref. [9] and alter Eq. (3) to the following

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{f} - \mathbf{f}_{\text{prior}}\|_2^2 \text{ such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \leq \epsilon \text{ and } \left\| \sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_p^p \leq \gamma, \quad (4)$$

which seeks the image $\mathbf{f}$ closest to a prior image $\mathbf{f}_{\text{prior}}$ while respecting constraints on the ℓ_p-norm of the GMI and data-error tolerance. We do not consider, here, the availability of a prior image and set $\mathbf{f}_{\text{prior}} = 0$, keeping this vector only for generality. Consider, first, the case of $p = 1$; the constraint on the GMI becomes a constraint directly on the image TV. Constrained minimization of image TV is known to encourage GMI sparsity. We do not directly minimize TV, rather we independently select parameters γ and ϵ. For sparsity-exploiting image reconstruction, both of these parameters are chosen to have small values: small ϵ forces tight agreement with the data, and small γ encourages GMI sparsity. We note that $\epsilon = 0$ corresponds to a data equality constraint, which may allow no solutions when inconsistencies are present in the data. For $p = 1$, the optimization problem in Eq. (4) is convex and the algorithm presented in Ref. [9] can be used directly to obtain the solution.

For this abstract, we are interested in developing an algorithm for $0 < p < 1$, where the GMI constraint becomes nonconvex. The issue then becomes how to solve Eq. (4) for $p < 1$, because the algorithm in Ref. [9] applies only to convex problems. The approach taken involves approximating Eq. (4) with a convex problem employing a weighted ℓ_1-norm:

$$\mathbf{f}^\circ = \arg\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{f} - \mathbf{f}_{\text{prior}}\|_2^2 \text{ such that } \|\mathbf{g}_{\text{data}} - X\mathbf{f}\|_2 \leq \epsilon \text{ and } \left\| \mathbf{w}\sqrt{(\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right\|_1 \leq \gamma, \quad (5)$$

where the GMI constraint involves only the ℓ_1-norm and a non-negative weight vector $\mathbf{w}$. For a given $\mathbf{w}$ this optimization problem is convex and can be solved efficiently using the algorithm in Ref. [9]. To attack the nonconvex problem, we employ a reweighting technique, where there are two loops: an inner loop where Eq. (5) is solved given parameters γ, ϵ, and weight vector $\mathbf{w}$, and an outer loop where the weight vector is adjusted based on the solution of the inner loop:

$$\mathbf{w} = \left(\sqrt{\eta + (\partial_x \mathbf{f})^2 + (\partial_y \mathbf{f})^2 + (\partial_z \mathbf{f})^2} \right)^{p-1}.$$

The parameter η is needed to prevent the singularity at voxels with zero GMI when $p < 1$. For all simulations in this abstract $\eta = 10^{-6}$. With a reweighting approach, an important question is how accurately does the intermediate weighted problem need to be solved in the inner loop so that overall convergence of the outer loop is attained. It turns out for the present reweighting scheme it suffices to have only one inner iteration. Thus, the complete algorithm is derived from the algorithm in Ref. [9], and the weights are recomputed at every iteration based on the current image estimate $\mathbf{f}$.

III. Results

To demonstrate the new image reconstruction algorithm, we perform two sets of experiments. In the first, we employ the algorithm on ideal, noiseless fan-beam CT data where it is possible to recover the exact image. With this ideal simulation, we demonstrate the potential for angular sampling reduction. In the second simulation, we apply the algorithm to circular, cone-beam CT projections with noise. The purpose of the latter simulation is to demonstrate that the algorithm can indeed by applied to 3D CT, and to illustrate the impact of the nonconvex algorithm on data inconsistency.

A. Ideal fan-beam CT simulation

For the 2D simulation we employ the breast phantom shown in Fig. 1. In the figure, the phantom GMI is also shown, which is seen to have many more zeros than the original phantom. It is this sparsity in the GMI, which we seek to exploit in order to reduce angular sampling. The phantom is discretized on a 128×128 pixel array, which is 20 cm on a side. Only the pixels within the largest inscribed circle are allowed to vary, and pixels outside this 20 cm diameter circle are fixed to zero. The fan-beam CT simulation models an X-ray source 40 cm from the isocenter, and a 80 cm source-to-detector distance.

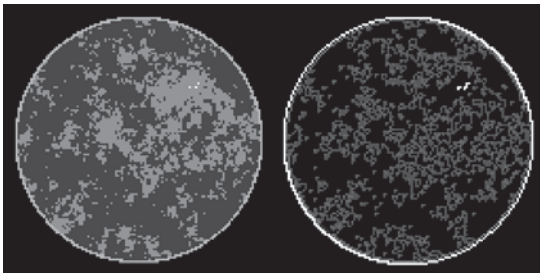


Fig. 1. Left: computerized breast phantom shown in a gray scale window [0.95, 1.25]. Right: gradient magnitude image (GMI), which has greater sparsity than the original phantom.

$p = 0.8$ $p = 1.0$ $p = 2.0$

Fig. 2. Reconstructed images for nonconvex $p = 0.8$, left column, compared with convex $p = 1.0$, middle column, and $p = 2.0$, right column. The number of views covering 360 degrees is 35, 30, and 25 for the top, middle, and bottom rows, respectively. The gray scale window is [0.95, 1.25].

The detector consists of 256 bins in a linear configuration, which is long enough to capture the projection of the 20 cm diameter pixel array. We consider only 360 degree scans, but allow the number of projections to vary.

To illustrate the potential of nonconvex optimization for sparsity-exploiting image reconstruction, we compare solutions of Eq. (4) for $p = 0.8$, $p = 1.0$, and $p = 2.0$. The latter two values lead to a convex problem, which can be solved with the algorithm in Ref. [9], and the first value leads to a nonconvex problem solved by the proposed reweighting algorithm using Eq. (5). For values $p = 1.0$ and $p = 2.0$, we have a direct convergence check, but for the nonconvex case we cannot claim to have found a global solution to Eq. (4). Instead, we can verify that Eq. (5) is solved for the weights $\mathbf{w}$ that are settled upon.

In applying the constraint-based optimization problem in Eq. (4), we need to specify two parameters ϵ and γ. The data used in this simulation are ideal, and accordingly we employ a tight data-error constraint and use a value for ϵ corresponding to an root-mean-square-error (RMSE) of 10^{-5}. For the image TV constraint we set γ to the value of the

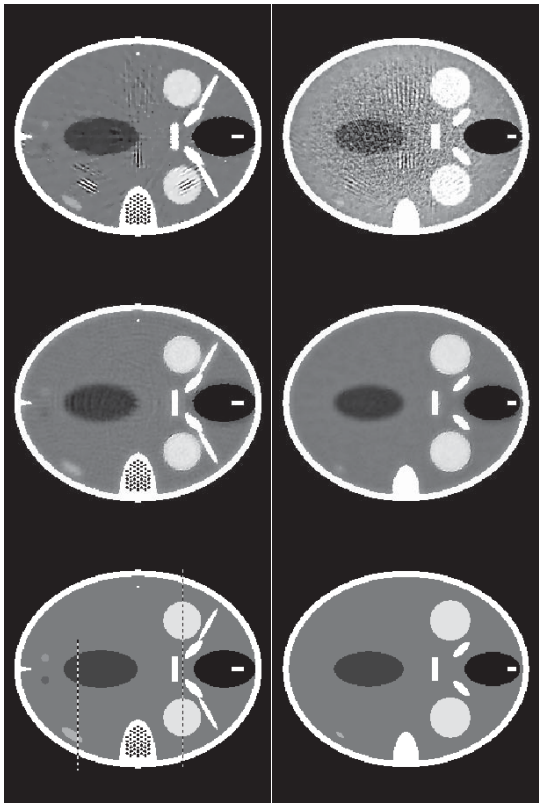


Fig. 3. Images reconstructed from noisy projections of the FORBILD head phantom. The rows show the results for $p = 1.0$, top, $p = 0.8$, middle, and the phantom, bottom in a gray scale window of [1.0425, 1.0625]. The first column shows the midplane, and the second column shows a transaxial plane near the top of the bony structure at the ear. The dashed lines in the phantom midplane slice indicate the locations of the profiles for Figs. 4 and 5.

ℓ_p-norm of the actual phantom GMI to the pth power. We note that in actual application, access to this information is unavailable and selection of γ would need to be based on different image quality metrics. Here, however, we are exploring the theoretical potential of the proposed algorithm.

Shown in Fig. 2, are image reconstruction results for 25, 30, and 35 simulated projections. The $p = 1.0$ case has some potential to reduce angular sampling by exploiting GMI sparsity. This is evident in the comparison with $p = 2.0$, which does not exploit GMI sparsity; the $p = 1.0$ results show visually accurate reconstruction for 35-view projection data while the $p = 2.0$ results do not show accurate reconstruction for any of the projection data sets. The nonconvex $p = 0.8$ results, however, extend the visually accurate reconstruction range down to 25-view projection data.

B. Circular cone-beam CT simulation with noisy projections

For the 3D circular cone-beam CT simulation, we scale up the problem approaching the scale of a realistic volume CT system, and we include noise on the CT projections. The phantom used for this simulation is the FORBILD head phantom, which has many low contrast objects, with gray level variations ranging from 0.25% to 1% of the phantom background, together with complex high-contrast bony structures. This phantom is quite challenging, because even minor streaks from the bony structures can interfere strongly with imaging the low-contrast objects.

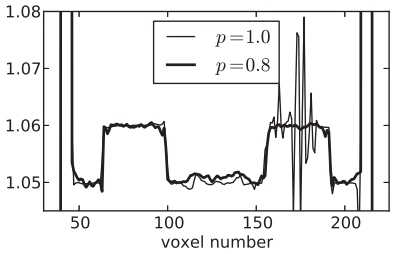


Fig. 4. Profile comparison corresponding to the images in Fig. 3 along a line in the midplane, through the eyes.

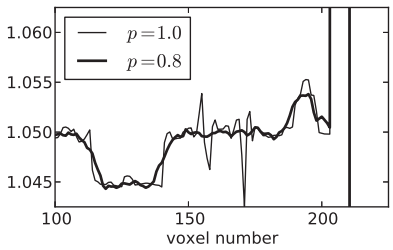


Fig. 5. Profile comparison corresponding to the images in Fig. 3 along a line in the midplane, through the ventricle and subdural hematoma.

The middle section of the head phantom is voxelized in a $256\times256\times32$ volume array, and the projection data simulate 100 projections onto a 512×80 bin flat-panel detector. Noise on the projections is modeled by employing independent 1D Gaussian distributions for each line-integration data value. The mean of each Gaussian distribution is the value of the corresponding line-integration over the phantom, and the standard deviation is taken to be 1% of this mean. The parameters of the simulation are such that it only makes sense to compare algorithms that exploit GMI sparsity, and accordingly we show results from Eq. (4) for $p = 0.8$ and $p = 1.0$. Larger p results in images that are heavily polluted with streak artifacts. For the constraint parameters, we employ an ϵ corresponding to an RMSE of 0.01, and for γ we use the value derived from the test phantom.

For the specified parameters, the image reconstruction results are shown in Fig. 3 together with corresponding slices in the phantom. The gray scale display window is 1% of the phantom complete dynamic range; and streak artifacts are difficult to avoid due to the rapidly oscillating bone structures near the ear at the bottom of the images. The results for $p = 1.0$, in the top row of the figure, show such streaks, even though this value for p does exploit GMI sparsity. The middle row shows results for the nonconvex case of $p = 0.8$, but the streak artifacts are nearly completely removed.

Inspection of the nonconvex results shows a rather interesting behavior in that the image regularization is highly non-uniform. The structures with the contrast of the eyes and greater ($\geq 1\%$ of phantom background) appear to have sharp edges, while the lower contrast structures are visible, yet, are blurred relative to the same structures in the $p = 1.0$ images. This visual impression is borne out quantitatively in vertical profile plots shown through the eyes, in Fig. 4, and through the ventricle and subdural hematoma, in Fig. 5. In the former profile, the nonconvex result has as sharp a transition at the eye border as the convex $p = 1.0$ result without the oscillations from the streaks. The latter lower contrast structures show fewer oscillation for the nonconvex result, but there is also a clear blurring as the transitions at the ventricle and hematoma borders are more gradual for $p = 0.8$ than for $p = 1.0$. This feature of the proposed nonconvex optimization can be understood from inspecting Eq. (5) where we see that the image TV term has a spatially dependent weighting. During the iteration of the nonconvex algorithm the weighting $\mathbf{w}$ evolves in such a way that less weight, and hence less smoothing, is applied to voxels with large gradient-magnitude.

IV. Summary

We have demonstrated GMI sparsity-exploiting image reconstruction by a nonconvex optimization algorithm. Under ideal conditions we have shown that the algorithm is capable of obtaining accurate image recovery with fewer projections than convex TV-based image reconstruction. The algorithm can also be applied to 3D cone-beam CT systems, and preliminary results indicate that the nonconvex algorithm can be effective in controlling steak artifacts resulting from a combination of projection view under-sampling and the presence of complex high-contrast structures.

V. Acknowledgments

This research was supported in part by the U.S. Department of Energy through the LANL/LDRD Program. This work is also part of the project CSI: Computational Science in Imaging, supported by grant 274-07-0065 from the Danish Research Council for Technology and Production Sciences. This work was also supported in part by NIH R01 grants CA158446, CA120540, and EB000225. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, 2008.

[2] Rick Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[3] Rick Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012.

[4] Emil Y. Sidky, Rick Chartrand, and Xiaochuan Pan, "Image reconstruction from few views by non-convex optimization," in *IEEE Nuclear Science Symposium Conference Record*, Honolulu, HI, 2007, pp. 3526 – 3530.

[5] E. Y. Sidky, X. Pan, I. S. Reiser, R. M. Nishikawa, R. H. Moore, and D. B. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Med. Phys.*, vol. 36, pp. 4920–4932, 2009.

[6] P. L. Combettes, "The foundations of set theoretic estimation," *IEEE Proc.*, vol. 81, pp. 182–208, 1993.

[7] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan, "Optimization-based reconstruction of sparse images from few-view projections," *Phys. Med. Biol.*, vol. 57, pp. 5245–5274, 2012.

[8] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, pp. 120–145, 2011.

[9] E. Y. Sidky, J. S. Jørgensen, and X. Pan, "First-order convex feasibility algorithms for X-ray CT," *Med. Phys.*, 2013, Accepted. Arxiv preprint (http://arxiv.org/abs/1209.1069).

Up–To–Date Prior Knowledge via Motion Correction for Low Dose Tomographic Fluoroscopy

Barbara Flach, Jan Kuntz, Marcus Brehm, Rolf Kueres, Sönke Bartling, and Marc Kachelrieß, *Member, IEEE*

***Abstract*—To provide 4D (=3D+time) volume data for interventional image guidance tomographic data acquisition during the whole intervention is necessary. But for the acceptance of the method, which is also called tomographic fluoroscopy or CT fluoroscopy, the patient dose level must be kept as low as in 2D+time fluoroscopic guidance, which is the standard imaging technique for today's intervention guidance.**

To achieve this goal a high quality prior volume acquired before intervention is necessary to guarantee time frames of high image quality during the intervention, despite the fact that these volumes are based on a very sparse angular sampling at very low dose. Depending on the type and duration of an intervention patient motion can become a problem. Consequently, the prior volume needs to be continuously updated.

We propose a prior that adapts itself by the combination of the two concepts registration and substitution. In the registration step a combination of affine and deformable registration adapts the prior to the current situation represented by a so–called target image. In the subsequent substitution step a forward projection of the deformed image yields virtual rawdata that are densely sampled in the angular direction. The latest measured projections are used to substitute the corresponding virtual projections. A reconstruction of these substituted data yields the adapted prior, the running prior.

Of course the success of the registration depends on the quality of the target image. Therefore we analyze what image quality is needed for the registration step to get acceptable results. For the evaluation of the running prior technique we used the head scan of a pig in vivo acquired by a prototype volume CT system.

We conclude that for preserving the image quality of the prior it is necessary to use more projections for the target image than used to reconstruct the temporal updates. However the interventional material can be also identified in the time frames if image quality is degraded by using only a reduced number of projections for reconstructing the target image.

***Index Terms*—Computed tomography (CT), interventional radiology, undersampled reconstructions, minimally–invasive interventions, C-arm**

I. Introduction

INTERVENTIONAL radiology comprises for example minimally–invasive procedures, stenting and aneurysm coiling where interventional material like guide wires, stents or coils are inserted into the patient's vessels [1]. Today, these procedures are guided via conventional (2D+time) fluoroscopy, which we will call projective fluoroscopy in the following. However, projective fluoroscopy only displays a superposition of the anatomy and therefore is not always well–suited to visualize complex structures and their spatial relationship. In these cases the interventionalist often applies trial–and–error approaches, e.g. to navigate an instrument through the vessels. This is time–consuming, increases patient dose and operation time, and leads in some cases to injuries of the vessel system.

Fig. 1. Volume rendering of two time frames of a stenting intervention reconstructed with PrIDICT. On the left side the stent is still closed and on the right side the stent is expanded.

Obviously, a continuous display of 3D volumes would be nice to have, a 3D+time modality which we will call tomographic fluoroscopy in the following. To keep the x–ray dose in tomographic fluoroscopy as low as that used in projective fluoroscopy it is necessary to obtain low dose tomographic data from a extremely low number of projections per (half or full) rotation of the imaging system.

Our prior (but preliminary) work indicated, that about 10 to 20 projections should be sufficient to update the volume like shown in figure 1 [2]–[4]. To achieve that, a dedicated reconstruction algorithm had to be applied that makes heavy use of the prior volume. Another option may be using the PICCS reconstruction algorithm of reference [5]. Existing reconstruction algorithms that do not make use of a prior scan, like the ASD–POCS algorithm [6] or the iTV algorithm [7], are not able to provide images of sufficient quality from this extremely high undersampling.

By now, all known methods provide satisfying results only in case of interventions without any patient motion. Our aim is to provide a robust method that addresses this problem and

Barbara Flach, Dr. Jan Kuntz, Marcus Brehm, Rolf Kueres, Dr. Sönke Bartling, and Prof. Dr. Marc Kachelrieß: Medical Physics in Radiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

Barbara Flach, Dr. Jan Kuntz, Marcus Brehm, and Prof. Dr. Marc Kachelrieß: Institute of Medical Physics, University of Erlangen–Nürnberg, 91052 Erlangen, Germany.

Dr. Sönke Bartling: Institute for Clinical Radiology and Nuclear Medicine, University Medical Center Mannheim, 68167 Mannheim, Germany

Corresponding author: barbara.flach@dkfz.de

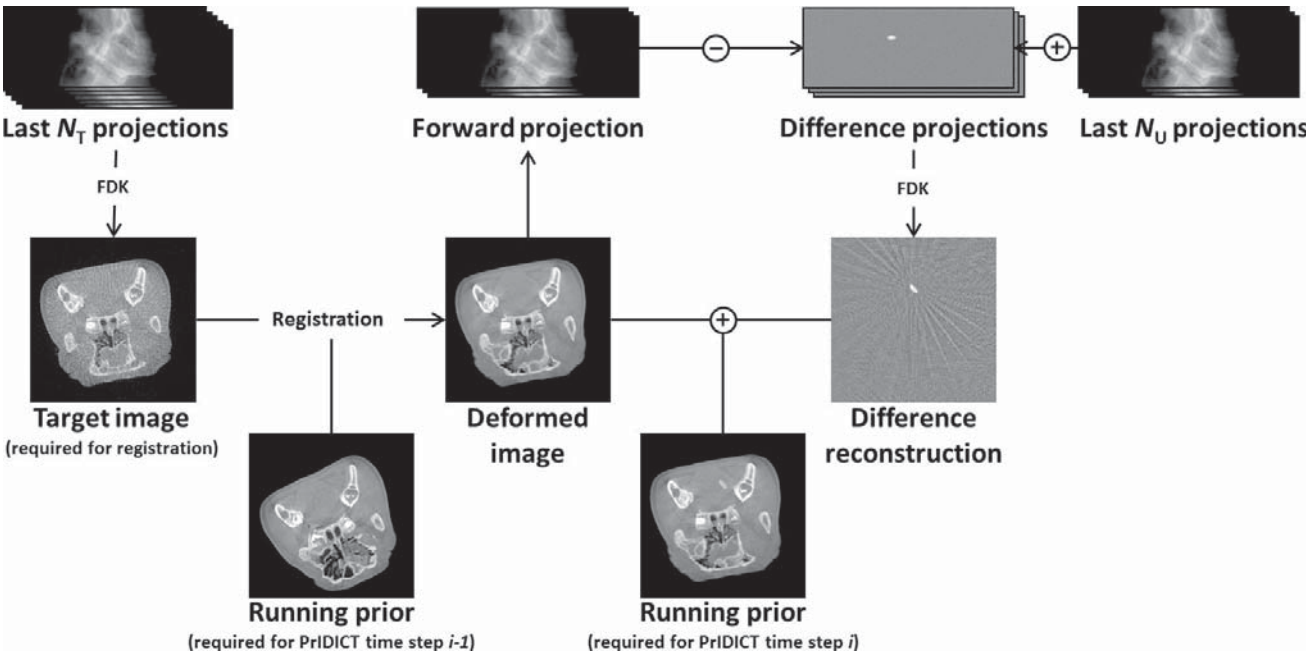


Fig. 2. Illustration of the updating process of the prior to yield a running prior which can be used in the PrIDICT algorithm. The last N_{U} projections are used to compute the time frame (in our experiments $N_{\mathrm{U}} = 15$. The last N_{T} projections are used to compute the target image (typically $N_{\mathrm{T}} \geq N_{\mathrm{U}}$).

allows for patient motion after the prior has been acquired. In particular it is required that the temporal resolution of the prior is high, corresponding to a minimum of projections, but with image quality being equivalent to a reconstruction from projections of a densely sampled rotation. We achieve this by continuously updating the prior in two steps. First via registration the prior is deformed to the current situation, which is represented by a so–called target image, and subsequently the new acquired projections are incorporated into the prior [8]. In praticular, the focus of this work lies on analyzing which image quality is required for the target image to get good image quality in the time frames.

II. Materials and Methods

A. Reconstruction Using Prior Data

We realize low–dose tomographic fluoroscopy using continuous CT scans with very few projections per rotation. Applying the prior image dynamic interventional computed tomography (PrIDICT) algorithm highly undersampled data of a half rotation is sufficient to update the volume [3], [4]. Due to the dose restrictions (that enforce sparse sampling) only about 10 to 20 projections are available. In our study we used $N_{\mathrm{U}} = 15$ projections for the calculation of each time frame. To obtain high quality images the information from a fully sampled ($N_{\mathrm{P}} = 600$ projections) prior scan is used during image reconstruction. For reconstruction the difference of the forward projection of the prior image and the rawdata of the intervention scan (the update scan) is computed and reconstructed. Ideally (in case of no motion) there is only interventional material visible. Because of the very sparse sampling the image quality is disturbed by streaks. To reduce these artifacts the L_0 norm is reduced by setting the values of insignificant voxels (= voxels with a low attenuation value) equal to zero. For display this image is added to the prior image, possibly using color coding. In addition a roadmap showing the vascular tree which is generated from a contrast–enhanced scan can be added to the prior, too. The resulting time frame shows the position and appearance of the interventional material in relation to the high quality 3D volume of the patient [3], [8].

B. Motion Correction

Up to here the algorithm works well in case of no or negligible patient motion. To deal with patient motion we introduce an additional step to adapt the high quality prior to the changes in patient and organ position and shape. Since the prior is adapted in each time step and this adapted prior serves as new prior for the PrIDICT algorithm in the corresponding time step, we call it running prior. For dose reasons only those projections shall be used to update the prior that are acquired anyway for the high temporal resolution intervention scan. The update is realized by a registration of the current prior to the target image, a Feldkamp-Davis-Kress (FDK) reconstruction [9] that uses the projections of some latest projections N_{T}. Here we investigated the results for parameters $N_{\mathrm{T}} = N_{\mathrm{U}} = 15$ and $N_{\mathrm{T}} = 120$. The registration algorithm consists of an affine registration [10] which is then refined by a deformable registration algorithm [11]. As a final step the deformed image is forward projected using a densely sampled geometry ($N_{\mathrm{P}} = 600$ projections, same number of projections as for the reconstruction of the prior image). From these newly obtained virtual rawdata the projections corresponding to projection angles of the latest N_{U} measured projections are substituted by the measured data. The resulting rawdata that consist of

$N_P - N_U$ forward projected and of N_U measured projections are reconstructed to yield the running prior. The adaptation of the prior is shown in figure 2. Here the replacement step is realized in a computationally more efficient but equivalent procedure. That is in detail by forward projecting only the N_U projections that shall be substituted, subtracting these projections from the N_U real measured projections and adding the reconstruction of this difference to the deformed image to obtain the running prior for this time step.

III. Experiments

We analyse the proposed method regarding the parameter N_T on the basis of a real dataset, the head scan of a pig in vivo. Since there exists no dedicated 3D+time interventional CT system so far we used a volume CT prototype, which is a flat detector mounted on a clinical gantry for continuous rotation which is necessary for the intervention scan (figure 3).

Fig. 3. Measurement setup with volume CT prototype and pig in vivo.

The scan for the prior image as well as the time frames is performed at 80 kV tube voltage and 50 mA tube current, but the scan for the time frames was pulsed resulting in a mean tube current of 18 mA. For the prior image we acquired 600 projections per full rotation at a rotation time of 20 seconds. The data for the intervention scan are acquired by a standard protocol with sampling rate of 30 frames per second, a gantry rotation time of 4 seconds. The total scan time was 30 seconds in this example. From these data we only used every fourth projection for our investigations, i.e. three out of four projections were completely ignored. This results in 15 used projections per half rotation. The intervention itself consisted of inserting a guide wire into the pig's carotid. Between prior and intervention scan a 30 mm shift of the pig's head and during the acquisition motion resulting from breathing occurred.

IV. Results

In figure 4 the target images, the running priors and the time frames for $N_T = 15$ and $N_T = 120$ at the beginning of the intervention (means one registration and replacement step) are compared. When using only $N_T = 15$ projections for the target image artifacts affect the registration process such that image parts are not correctly adapted like the two regions pointed to by arrows in figure 4. In addition the result with $N_T = 15$ shows a wobbly bodyoutline for example in the encircled area. This is not the case for $N_T = 120$.

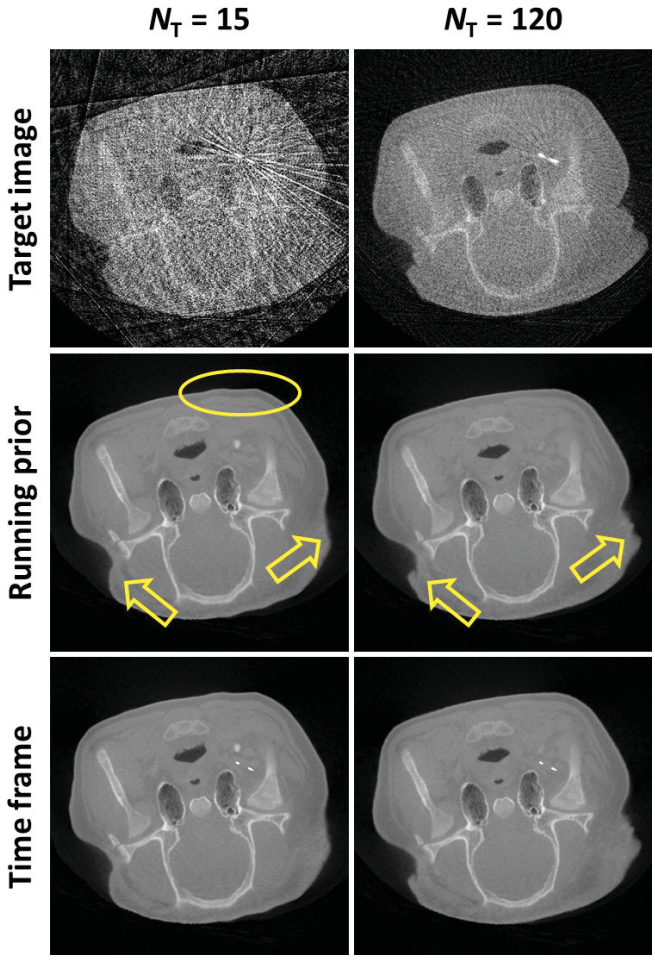


Fig. 4. Comparison of target image, running prior and time frame based on $N_T = 15$ and $N_T = 120$ at the beginning of the intervention. The images are displayed at a gray scale window $C/W = 0$ HU/2000 HU.

At the end of the intervention the running prior with $N_T = 120$ shows good image quality. In contrast the running prior with $N_T = 15$ looks quite smooth like shown in figure 5. This is a consequence of more small errors in the vector field (resulting from the quite poor image quality with $N_T = 15$) and the subsequent interpolation within the deformation.

The essential in interventional image guidance is that the wire can be distinguished and this of course at the correct position. This is possible in both cases like demonstrated in figure 6.

V. Discussion

The running prior technique guarantees significantly increased robustness compared to using a static prior. But the exactness of the results strongly depends on the result of the registration. It is a trade–off between a good image quality of the target image (high N_T), such that the registration result

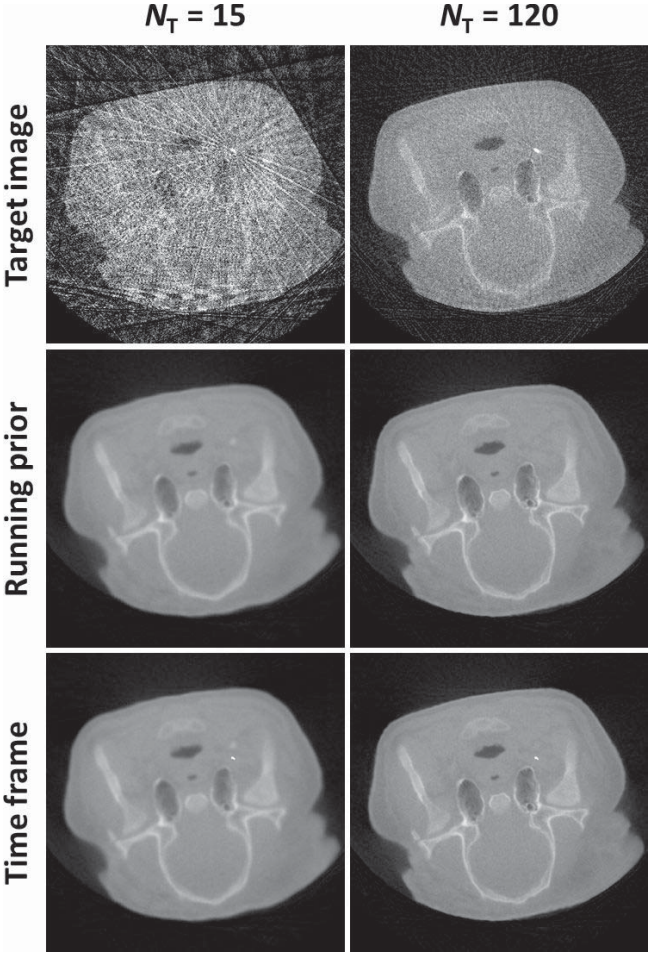


Fig. 5. Comparison of target image, running prior and time frame based on $N_T = 15$ and $N_T = 120$ at the end of the intervention. The images are displayed at a gray scale window $C/W = 0$ HU$/2000$ HU.

is not affected by the artifacts included in the target image, and on the other side the actuality of the target image (lower N_T). Due to our results we conclude that $N_T = N_U = 15$ projections provide satisfying results for interventional image guidance. The interventional material can be displayed at the correct position. But for keeping the image quality of the running prior almost at the same level as that of the static prior acquired before intervention, it is necessary to use some more projections, although not necessarily $N_T = 120$. By this good image quality is possible at the expense of a temporally slightly delayed running prior.

ACKNOWLEDGMENT

This study was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant KA 1678/6-1. The high performance compute hardware was provided by the Universitätsbund Erlangen-Nürnberg e.V., Erlangen, Germany. Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

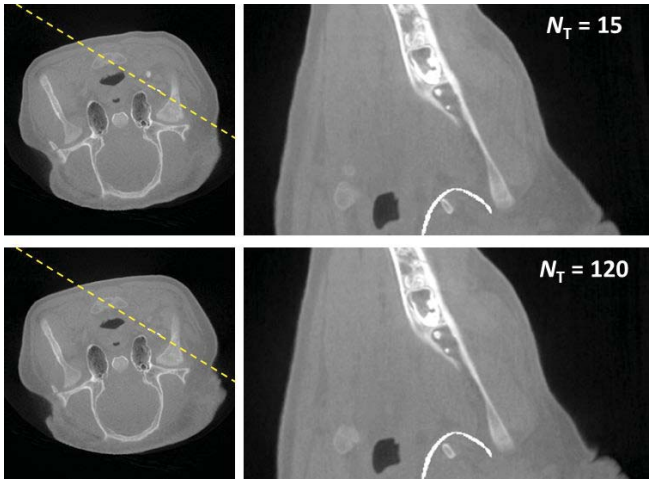


Fig. 6. Comparison of the identification of interventional material based on $N_T = 15$ and $N_T = 120$ projections for the target image. In the left column a transversal slice of the time frame is shown. In the right column the slice corresponging to the yellow line through all z-positions is shown. In the upper row for $N_T = 15$ and in the lower row for $N_T = 120$. The images are displayed at a gray scale window $C/W = 0$ HU$/2000$ HU.

REFERENCES

[1] T. Struffert and A. Dörfler, "Flachdetektor–CT in der diagnostischen und interventionellen Neuroradiologie," *Der Radiologe*, vol. 49, no. 9, pp. 820–828, Sep. 2009.

[2] J. Kuntz, M. Brehm, M. Kachelrieß, and S. Bartling, "Towards 4D intervention guidance using compressed sensing," *Proceedings of the 11th International Meeting on Fully 3D Image Reconstruction*, pp. 347–350, Jul. 2011.

[3] J. Kuntz, R. Gupta, S. Schönberg, W. Semmler, M. Kachelrieß, and S. Bartling, "Real–time X–ray–based 4D image guidance of minimally invasive interventions," *European Radiology*, Jan. 2013.

[4] J. Kuntz, B. Flach, R. Kueres, W. Semmler, M. Kachelrieß, and S. Bartling, "Constrained reconstructions for 4D intervention guidance," *Accepted for publication at Phys. Med. Biol.*

[5] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 25, no. 2, pp. 660–663, Feb. 2008.

[6] E. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few–views and limited–angle data in divergent–beam CT," *Journal of the Optical Society of America*, vol. 25, no. 7, pp. 1772–1782, Jul. 2008.

[7] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, pp. 1545–1561, Mar. 2011.

[8] B. Flach, J. Kuntz, M. Brehm, S. Bartling, and M. Kachelrieß, "Running prior for patient motion correction in low–dose 3D+time interventional flat detector CT," *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pp. 2395–2401, Oct. 2012.

[9] L. Feldkamp, L. Davis, and J. Kress, "Practical Cone–Beam Algorithm," *Journal of the Optical Society of America*, vol. 1, no. 6, pp. 612–619, Jun. 1984.

[10] P. Viola and W. I. Wells, "Alignment by Maximization of Mutual Information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[11] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, Sep. 1998.

Nested Loop Algorithm for Parallel Model Based Iterative Reconstruction

Zhou Yu, Lin Fu, Debashish Pal, Jean-Baptiste Thibault, Charles A. Bouman and Ken D. Sauer

***Abstract*— Model based iterative reconstruction (MBIR) algorithms have been used in clinical studies to allow significant dose reduction in CT scans while maintaining the diagnostic image quality. Simultaneous-update algorithms, which can take advantage of massively parallel computer architectures, are promising to significantly improve the speed of MBIR. To achieve this goal, we also need to improve the convergence speed of these algorithms. In this paper, we propose a fast converging simultaneous-update algorithm using a nested loop structure. Preliminary experimental results show that the proposed algorithm has faster convergence speed compared to algorithms such as conjugate gradient and preconditioned conjugate gradient methods.**

***Index Terms*— Computed tomography, iterative reconstruction, nested loop, preconditioner.**

I. Introduction

Recent applications of model based iterative reconstruction (MBIR) algorithms to medical CT have demonstrated significant improvement in image quality by increasing resolution as well as reducing noise and artifacts [1], [2]. Clinical studies also show that MBIR algorithms can be used as a tool to allow significant dose reduction in CT scans while maintaining diagnostic image quality [3]. With ever advancing computing technologies, massively parallel architectures have emerged, such as the newest multi-core CPUs and GPUs. These new hardware technologies bring promise to significantly speed up the MBIR algorithms [4]. Taking advantage of these new technologies requires developing algorithms that are highly parallel and yet have fast convergence properties. Simultaneous-update algorithms, such as variations of expectation maximization (EM) [5], conjugate gradients (CG) [6], and ordered subsets (OS) [7], are attractive since they are easier to map on to highly parallel computer architectures to reduce per iteration computational cost. However, compared to sequential-update algorithms such as iterative coordinate descent (ICD) [8], [9], simultaneous-update algorithms tend to require many more iterations to converge. Therefore, it is critical to speed up the convergence of simultaneous-update algorithms.

In this paper, we propose a nested loop framework to accelerate the convergence of simultaneous-update algorithms. Our method is composed of inner and outer loop iterations. In each outer loop iteration, we create a local approximation to the cost function. The approximate problem is then solved by inner loop iterations with relatively low computational cost. The inner loop solution is used to compute an update direction for the outer loop. The outer loop then computes an optimal step size so that it guarantees the cost function will decrease monotonically. Similar nested loop algorithms have been explored in PET reconstruction problems [10]–[12]. In the CT reconstruction problem, we propose to construct the inner loop problem using an image space approximation to the Hessian matrix of the original cost function.

The project described was partly supported by Grant No. 1-R01-HL-098686 from NIH and its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

Zhou Yu, Debashish Pal and Jean-Baptiste Thibault are with GE Healthcare Technologies, 3000 N Grandview Bvd, W-1180, Waukesha, WI 53188.

Lin Fu is with GE Global Research Center

Charles Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-0501.

Ken Sauer is with the Department of Electrical Engineering, 275 Fitzpatrick, University of Notre Dame, Notre Dame, IN 46556-5637.

II. Method

A. Objective Function

MBIR algorithms typically work by first forming an objective function which incorporates an accurate system model [13], statistical noise model [1] and prior model [14]. The image is then reconstructed by computing an estimate which minimizes the resulting objective function.

Let x denote the image and y denote the measurement data. We consider both x and y as random vectors, and our goal is to reconstruct the image by computing the maximum *a posteriori* (MAP) estimate given by

$$x^* = \arg\min f(x) \tag{1}$$

$$f(x) = \{\frac{1}{2}J(x,y) + \Phi(x)\} \tag{2}$$

where $J(x,y)$ is the log likelihood term that penalizes the inconsistency between the image and the measurement, $\Phi(x)$ is the negative log of the prior distribution that penalizes the noises in the image. One example of $J(x,y)$ is in quadratic form

$$J(x,y) = ||Ax - y||^2_W \tag{3}$$

where A is the system matrix, W is a diagonal weighting matrix. The i^{th} diagonal entry of the matrix W, denoted by w_i, is typically chosen to be approximately inversely proportional to the estimate of the variance in the measurement y_i [1], [9]. We will consider the data mismatch term in (3) to illustrate the algorithm framework in this paper. However, the proposed algorithm can also be applied to other forms of data mismatch terms, such as the Poisson log likelihood function, as long as $f(x)$ remains strictly convex.

B. Nested Loop Algorithm

The idea in this paper is to create a sequence of sub-problems that optimize simpler approximate cost functions, while still guaranteeing that the solution of the sub-problems will converge to the solution of the original cost function. To achieve this, we propose a nested loop framework. Let n be

the outer loop iteration index, and $x^{(n)}$ be the image estimate after the n^{th} iteration. In each outer loop iteration, we first create a local approximate cost function, $h^{(n)}(x)$, which must satisfy,

$$\nabla h^{(n)}(x^{(n-1)}) = \nabla f^{(n)}(x^{(n-1)}) \quad (4)$$

We then minimize $h^{(n)}$ using inner loop iterations. If we update the image directly using the inner loop solution, this does not necessarily guarantee convergence. Instead, we use the solution of the sub-problem to compute an update direction, and then solve a 1D optimization problem to determine the update step size. Since the cost function is minimized along the search direction, it is guaranteed to decrease monotonically with every outer loop iteration.

In the following, we propose a method to apply the nested loop framework to CT iterative reconstruction problem. First, we need to derive the approximate cost function used in each outer loop. We can rewrite the cost function in (2) and (3) as

$$f(x) = ||x - x^{(n-1)}||^2_{A^tWA} + x^t\Theta^{(n-1)} + \Phi(x) + c^{(n-1)}, \quad (5)$$

where $\Theta^{(n-1)} = A^tW(Ax^{(n-1)} - y)$ and $c^{(n-1)}$ is a constant. In CT iterative reconstruction, the most expensive computation components in each iteration are typically the forward projection and the back projection, i.e. A and A^t. Therefore, we create the approximate cost function by replacing A^tWA in (5) with a simpler operator M, i.e.

$$h^{(n)}(x) = ||x - x^{(n-1)}||^2_M + x^t\Theta^{(n-1)} + \Phi(x) + c^{(n-1)} \quad (6)$$

Written in this form, it is easy to verify that the condition in (4) holds for the proposed $h^{(n)}(x)$. Notice that, the regularization term in $h^{(n)}(x)$ is also calculated exactly. The only approximation in $h^{(n)}(x)$ is in the second derivative of the cost function by replacing A^tWA with M.

Designing M appropriately requires balancing between two objectives. First, M needs to be a close approximation of A^tWA. Second, M must be easy to pre-compute so the matrix vector multiplication can be computed at low cost. In this paper, we use the approximation to A^tWA operator proposed by Fessler and Booth in [15], that is,

$$A^tWA \approx DKD, \quad (7)$$

where D is a diagonal matrix, with the j^{th} diagonal element to be $d_i = \sqrt{\frac{\sum_j a_{ij}^2 w_j}{\sum_j a_{ij}^2}}$, and K is a circulant matrix approximation to the A^tA operator. Notice that the proposed operator is a pure image space operator. It is very easy to compute since it only requires image scaling and filtering, which can be efficiently implemented with fast Fourier transforms (FFT).

Second, we solve the minimization problem of $h^{(n)}$ iteratively using inner loops. The inner loop problem described in (6) is similar to an image space denoising problem. One can solve this problem with a simple gradient based method, such as gradient descent. Here, we can also accelerate the inner loop convergence using an image space preconditioner, such as Fourier based preconditioners proposed in [15] and [16].

Finally, we update the image based on the inner loop solution. Let $\tilde{x}^{(n)}$ denote the solution of the inner loop. We compute the update direction using $u^{(n)} = \tilde{x}^{(n)} - x^{(n-1)}$, and then compute the step size $\beta^{(n)}$. The step size is computed to minimize the cost along the update direction, that is,

$$\beta^{(n)} = \arg\min_{\beta} f\left(x^{(n-1)} + \beta u^{(n)}\right) \quad (8)$$

This way, we can guarantee $f(x^{(n)}) < f(x^{(n-1)})$. We can compute an approximate solution to (8) using a closed form formula similar to the one used in [15]. Since the step size formula is derived using a surrogate function, the monotonicity of sequence $\{f(x^{(n)})\}$ is still guaranteed.

Fig. 1 shows the pseudo code of the proposed algorithm. In each outer loop, we first formulate an approximate cost function, $h^{(n)}(x)$, which is then optimized by the inner loop iterations in line 5. By eliminating A and A^t operation in $h^{(n)}$, each inner loop has very low computational cost. In line 6, the result of the inner loop is used to compute the update direction, followed by the step size calculation in line 7.

The outer loop algorithm is similar to a gradient descent algorithm except the search direction is computed from the inner loop solution rather than the gradient direction. We can easily generalize this algorithm by using other gradient based method in the outer loop such as conjugate gradient, etc.

1: $x \leftarrow$ FBP reconstruction
2: $\Theta \leftarrow A'W(Ax - y)$
3: **repeat**
4: $\quad x_{prev} \leftarrow x$
5: $\quad x \leftarrow \arg\min_v ||v - x_{prev}||^2_M + v^t\Theta + \Phi(v)$
6: $\quad u \leftarrow x - x_{prev}$
7: $\quad \beta \leftarrow \arg\min_\beta f\left(x_{prev} + \beta u\right)$
8: $\quad x \leftarrow x_{prev} + \beta u$
9: $\quad \Theta \leftarrow \Theta + \beta A'WAu$
10: **until** x is converged

Fig. 1. Pseudo code of one example of nested loop algorithm. In each outer loop, we first formulate an approximate cost function. Line 5 is solved iteratively using the inner loop iterations. The result of the inner loop is used to compute an update direction in line 6. Finally, the line search step in line 7 guarantees the monotonicity of the cost function.

III. Experimental Results

In this section, we compare the convergence speed of Nested Loop algorithm to CG and two variations of Preconditioned CG algorithms. Here, the nested loop (NL) algorithm and PCG algorithms use the same approximation to A^tWA with image space operator DKD, in which K is a shift invariant filter with frequency response of the form $K(f) = \frac{1}{|f|+c}$, where $f \in [-0.5, 0.5]$. The difference among the three methods lies in the modeling of the regularization function. In the method labeled as PCG-P1, the Hessian of the regularization term is ignored. Therefore, its preconditioner is given by

$$P_1 = D^{-1}K^{-1}D^{-1}. \quad (9)$$

In the method labeled as PCG-P2, we model the Hessian of the regularization term as DRD, where R has the frequency response

$$R(f) = 1 - \cos(2\pi f). \quad (10)$$

In this case, the preconditioner is given by

$$P_2 = D^{-1}(K + \alpha R)^{-1} D^{-1}. \quad (11)$$

Finally, in the nested loop algorithm, we compute the regularization term exactly in the inner loop problem. The inner loop is solved using 10 iterations of PCG, where we use P_1 as the preconditioner.

Fig. 2 shows the Fourier transform of the preconditioner kernel used in the experiment. Notice that, by modeling the regularization term in the cost function, the high frequency gain in the preconditioner P_2 is suppressed.

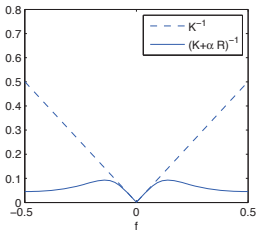


Fig. 2. This figure shows the Fourier transform of the preconditioner kernel used in the experiment, K^{-1} and $(K + \alpha R)^{-1}$, where $\alpha = 10$

The data we use to test the algorithm is a low dose axial scan of the GE performance phantom shown in Fig. 3. In this phantom, wires and resolution bars are used to measure the spatial resolution of the reconstruction, and a uniform region provides the noise measurements.

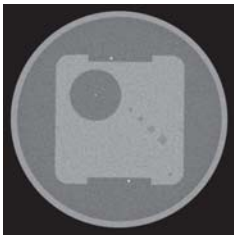


Fig. 3. This figure shows GE performance phantom used in the experiment. We use the wire in the phantom to measure in-plane resolution and the uniform region to measure the noise standard deviation

In Fig. 4, the results of various algorithms are compared against a reference image computed from a sufficiently converged reference NH-ICD [9] algorithm. In (a), we compute the root mean squared difference (RMSD) in the ROI volume between the current image and the reference image. Fig. 4 (b) and (c) characterize the convergence speed of high frequency components for each algorithm. In (b), we use the wire in the image to measure the 50 percent MTF after each iteration, and then plot the MTF convergence curve. The horizontal dashed line shows the MTF achieved by reference NH-ICD algorithm sufficiently converged after 20 iterations. In (c), we measure the noise standard deviation in a uniform ROI after each iteration. The horizontal line shows the noise level in a the reference NH-ICD reconstruction. Comparing the convergence plots, we find a consistent trend: As expected, preconditioning significantly accelerates the convergence speed of CG. By including the regularization term in the preconditioner design (P_2), we can further speed up the convergence. Finally, the NL method has the fastest convergence speed, which can be attributed to its ability to compute the regularization function exactly in the inner loops.

As shown in (c), in the uniform area, the PCG-P2 algorithm has a similar convergence speed to the NL algorithm. However, in (b), the NL method appears to be much faster than the PCG P_2 method. This is probably because, in the PCG-P2 method, we choose the parameter α to match the regularization strength in the uniform area. Since we use an edge-preserving regularization function, the same parameter can be too strong in the area around the wire causing slower resolution recovery. On the other hand, since the NL algorithm computes regularization term exactly, it shows consistently fast convergence speed in both the uniform area and around the wire without the need to choose parameters to optimize the convergence behavior.

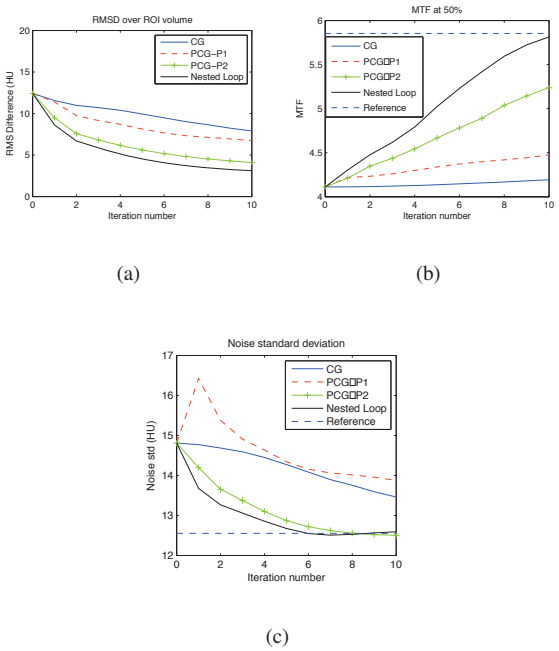
Fig. 5 shows the reconstructed image from a body scan data, in which (a) shows the denoised FBP image used as initial condition for the iterative reconstruction, (b) and (c) shows the image after 3 and 10 iterations of the nested loop algorithm, and (d) shows the fully converged image generated using 20 iterations of NH-ICD algorithm. The figure shows the image resolution improves very quickly using the proposed algorithm, and reaches the same solution as the NH-ICD algorithm.

Finally, let us comment on the computational cost of each algorithm. Typically, forward and back projection are the most computationally expensive components of each iteration. The cost of applying preconditioners (additional FFTs) is relatively low. Therefore, the per iteration cost of the PCG algorithm and the CG algorithm are very similar. The nested loop algorithm has slightly higher computational cost mainly due to multiple FFTs per inner loop iteration. However, in general, the inner loop cost may still be low compared to the projector costs. In practice, one can adjust the number of inner loop iterations to balance convergence speed and per iteration computation cost.

IV. Discussion and Future Work

The idea of using an approximate cost function is also used in the optimization transfer techniques. In these methods, one typically designs a surrogate function that satisfies equation (4), and upper bounds $f(x)$ for all x. However, to satisfy the upper bound condition, $h^{(n)}(x)$ tends to have very large curvature, and therefore, the update steps tend to be very small. In our approach, we replace the upper bound condition with a line search step, giving us more flexibility to explore different choices of $h^{(n)}(x)$.

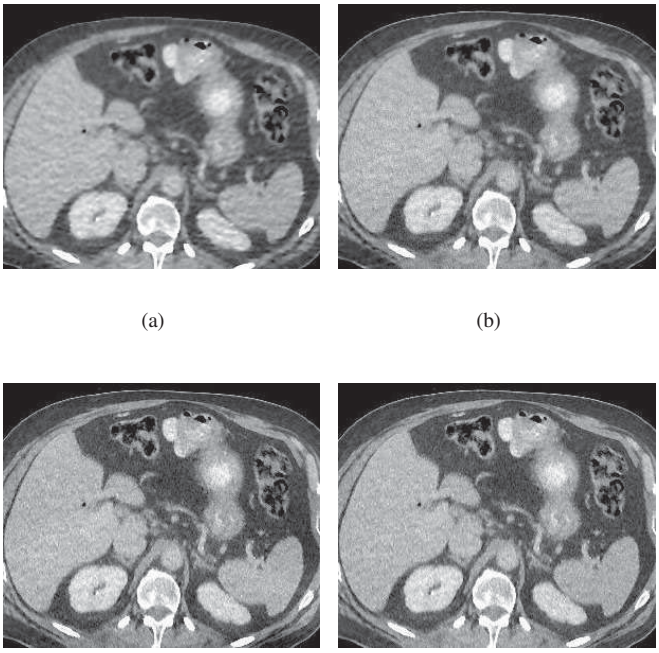
The idea to approximate $A^t W A$ has been well explored in preconditioner methods. In these methods, the idea is to design a matrix operator P directly to approximate the inverse of the Hessian $H^{-1}(x)$. However, in our approach, we only attempt to approximate the forward Hessian matrix. In general, approximating $H(x)$ instead of $H^{-1}(x)$ has two benefits. First, we can compute part of the cost function, in our case, the regularization term, exactly. In the typical preconditioner methods, the Hessian matrix of the regularization function is either completely ignored, or approximated. Using the same approximation to $A^t W A$, the proposed method outperforms



(a) (b)

(c)

Fig. 4. This figure shows convergence curves comparing the results of various algorithms. In (a), we show the root mean squared difference (RMSD) between the image and the fully converged reference image. In (b), we use the wire in the image to measure the 50 percent MTF after each iteration, and then plot the MTF convergence curve. The horizontal dashed line shows the MTF achieved by ICD algorithm after 20 iterations. In (c), we measure the noise standard deviation in a uniform ROI after each iteration. The horizontal line shows the noise level in a fully converged ICD reconstruction. All convergence curves show that the nested loop algorithm has the fastest convergence speed of all methods considered here.



(a) (b)

(c) (d)

Fig. 5. This figure shows the reconstructed image of a body scan, in which (a) shows the FBP initial image, (b) and (c) shows the image after 3 and 10 iterations, and (d) shows the fully converged image generated using NH-ICD. The figure shows that the image resolution improves very quickly using the proposed algorithm.

the preconditioner based methods mostly because it models the regularization function exactly. Second, it provides us the flexibility to use more sophisticated and more accurate models of the Hessian. In this paper, we construct the inner-loop algorithm using well-known approximations. In future work, we will further explore the flexibility of the algorithm to design a more sophisticated approximation of the $A^t W A$ operator.

REFERENCES

[1] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[2] A.Ziegler, Th.Köhler, and R.Proksa, "Noise and resolution in images reconstructed with FBP and OSC algorithms for CT," *Med. Phys.*, vol. 34, no. 2, pp. 585–598, 2007.

[3] G. Yadav, S. Kulkarni, Z. R. Colon, J. Thibault, and J. Hsieh, "Dose reduction and image quality benefits using model based iterative reconstruction technique for computed tomography," in *Fifty-second annual meeting of the American Association of Physicists in Medicine*, 2010.

[4] K. Mueller and R. Yagel, "Fast implementations of algebraic methods for three-dimensional reconstruction from cone-beam data," *IEEE Trans. on Medical Imaging*, vol. 18, no. 6, pp. 538–548, 1999.

[5] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, no. 2, pp. 113–122, October 1982.

[6] E. Ü. Mumcuoğlu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[7] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 601–609, December 1994.

[8] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[9] Z. Yu, J.-B. Thibault, C. Bouman, K. Sauer, and J. Hsieh, "Fast model-based X-Ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. on Image Processing*, vol. 20, no. 1, pp. 161–175, 2011.

[10] L. Fu, "Residual correction algorithms for statistical image reconstructionin positron emission tomography," Ph.D. Thesis, University of California, Davis, Davis, Ca, Feb 2010.

[11] G. Wang and J. Qi, "Acceleration of the direct reconstruction of linear parametric images using nested algorithms," *Physics in Medicine and Biology*, vol. 55, no. 5, 2010.

[12] I. Hong, Z. Burbar, C. Michel, and R. Leahy, "Ultrafast preconditioned conjugate gradient algorithm for fully 3D PET reconstruction," in *IEEE Medical Imaging Conference*, Orlando, FL, Oct. 2010.

[13] B. DeMan and S. Basu, "Distance-driven projection and backprojection in three-dimensions," *Physics in Medicine and Biology*, vol. 49, pp. 2463–2475, 2004.

[14] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[15] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. on Image Processing*, vol. 8, no. 5, pp. 688–99, 1999.

[16] L. Fu, B. DeMan, K. Zeng, T. M. Benson, Z. Yu, G. Cao, and J.-B. Thibault, "A preliminary investigation of 3D preconditioned conjugate gradient reconstruction for cone-beam CT," in *SPIE Medical Imaging Conference*, 2012, pp. 8313–134.

Statistical Reconstruction for Soft Tissue Imaging with Low Dose C-arm Cone-Beam CT

Adam S. Wang, J. Webster Stayman, Yoshito Otake, Gerhard Kleinszig, Sebastian Vogt,
A. Jay Khanna, Ziya L. Gokaslan, and Jeffrey H. Siewerdsen

***Abstract*—C-arm cone-beam CT (CBCT) is entering a growing scope of applications in intraoperative imaging but typically exhibits limited soft-tissue imaging capability and is largely constrained to high-contrast imaging tasks. Statistical iterative reconstruction techniques offer major advances in image quality and considerable promise for low-dose soft-tissue imaging. This work adapts penalized likelihood (PL) reconstruction approaches facilitated by high-speed computing to C-arm CBCT and investigates performance in low-dose imaging of low-contrast (<100 HU) tasks pertinent to soft-tissue surgical guidance. Fair comparison of image quality performance in 3D filtered backprojection (FBP) and PL exercised careful matching of 3D spatial resolution (viz., matching the edge spread function across a low-contrast sphere) since each approach has parameters allowing distinct tradeoffs of noise and resolution.**

An anthropomorphic abdominal phantom with various tissue-equivalent inserts was used to quantify contrast-to-noise / resolution tradeoffs in low-contrast structures when imaged using a mobile C-arm for CBCT. These comparisons of FBP and PL reconstructions across a range of low-dose protocols provided insight on low-dose limits, and sparse acquisitions were also considered as a method to reduce dose and data size. Statistical reconstruction increased soft-tissue image quality through reduction of noise and artifacts (e.g., 2-3 fold increase in CNR at equivalent spatial resolution), enabling a corresponding dose reduction wherein not only was the dose lower but also the image quality improved relative to FBP. The optimal reconstruction parameters were translated to imaging a cadaveric torso, where improved visualization of soft-tissue structures was confirmed. The advances in low-dose soft-tissue image quality offered by statistical reconstruction demonstrate promise for intraoperative C-arm CBCT to overcome conventional tradeoffs in noise, resolution, and dose.

***Index Terms*—Image Quality, Cone-Beam CT, Statistical Reconstruction, Imaging Task, Radiation Dose, Image Guidance**

I. Introduction

Intraoperative imaging using mobile or fixed-room C-arms provides real-time 2D imaging (fluoroscopy) and volumetric 3D imaging (cone-beam CT, CBCT) for image-guided procedures. While 3D imaging on C-arms has become more prevalent, conventional filtered back-projection (FBP) image reconstruction exhibits limited image quality, with low-contrast visibility being significantly less than diagnostic CT due to higher levels of noise and artifacts arising from scatter, truncation, and view sampling [1, 2]. Thus, low-contrast soft tissues requiring contrast resolution at the level of 10s of HU are largely beyond the current capabilities of CBCT using FBP but may be enabled by iterative reconstruction. Additionally, the need to minimize radiation dose to both the patient and surgical staff motivates the development of low-dose imaging techniques. Enabling high-quality, low-dose C-arm CBCT would therefore provide physicians with new capabilities for soft-tissue surgical guidance, such as more precise targeting of tumors and avoidance of adjacent critical anatomy.

This work leverages advances in statistical iterative reconstruction and computational hardware to enable better visualization of low-contrast soft-tissues. Advanced reconstruction methods such as penalized likelihood (PL) have increasingly attracted attention to methods that account for factors in the imaging chain to improve image quality and enable low-dose protocols. While more computationally expensive than FBP, parallelizable implementations on high-speed computer architecture continues to accelerate statistical reconstruction methods to meet the demands of clinical workflow. We investigate the potential for PL reconstruction to improve soft-tissue visibility in mobile C-arm CBCT and enable lower dose imaging protocols while maintaining image quality. In addition to lowering the exposure per projection (tube current), sparse sampling of projections is readily accommodated by C-arm CBCT and is investigated not only as a method to reduce dose but also to potentially enable faster reconstructions due to smaller data sets. Methods for fair comparison of FBP and PL, including noise/resolution tradeoffs at matched spatial resolution as measured by the edge spread function across low-contrast spheres, provided quantitative evaluation of low-dose reconstructions in an anthropomorphic phantom and helped to specify reconstruction parameters to be applied in imaging a cadaver for soft-tissue imaging in the abdomen.

II. Methods

A. Experimental Setup

A prototype mobile C-arm system with a flat-panel detector (PaxScan 3030+, Varian) was used for volumetric imaging of a phantom and cadaver (Fig. 1). The system has a 16×16×16 cm^3 field of view and angular range of 178°. A typical low-dose CBCT acquisition collected 200 projections [768×768 pixels (2×2 mode)] at 100 kVp, 60 mAs, corresponding to 1.3 mGy dose (at the center of a 32 cm body phantom) compared to a typical dose in diagnostic abdominal CT in the range of 10-25 mGy [3]. All images were reconstructed with isotropic 0.6×0.6×0.6 mm^3 voxels. Sparse acquisitions (e.g., undersampling the projections by 2×) were considered utilizing every other projection for reconstruction.

This work is supported by academic-industry partnership with Siemens Healthcare, XP Division.

A. S. Wang, J. W. Stayman, Y. Otake, and J. H. Siewerdsen are with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA (e-mail: {adamwang, web.stayman, otake, jeff.siewerdsen}@jhu.edu).

G. Kleinszig and S. Vogt are with Siemens Healthcare, Erlangen, Germany (e-mail: {gerhard.kleinszig, sebastian.vogt}@siemens.com).

A. J. Khanna is with the Department of Orthopaedic Surgery, Johns Hopkins University, Baltimore, MD 21205 USA (e-mail: akhanna1@jhmi.edu).

Z. L. Gokaslan is with the Department of Neurosurgery, Johns Hopkins University, Baltimore, MD 21205 USA (e-mail: zgokasl1@jhmi.edu).

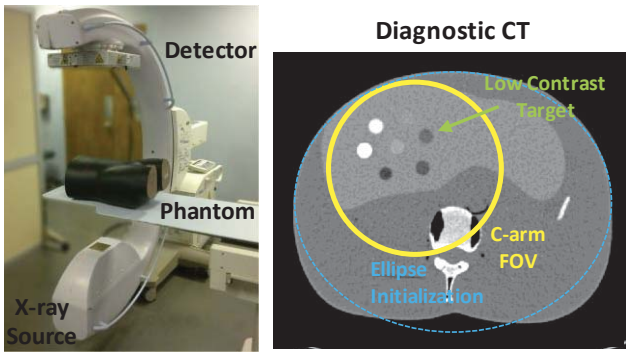


Fig. 1. Mobile C-arm for intraoperative CBCT and the anthropomorphic abdomen phantom containing a simulated liver and low-contrast inserts.

An anthropomorphic abdomen phantom (Fig. 1) was used to quantify the imaging performance of FBP and PL reconstructions at different dose levels, ranging from 15 to 120 mAs (0.3 to 2.6 mGy) at 100 kVp. The simulated liver contained 10 mm diameter spheres of varying contrast, and the low-contrast sphere (approximately -70 HU) was selected as the target for quantitative analysis. The same dose levels and reconstruction parameters were used in imaging a fresh cadaver torso, which presented realistic soft-tissue visualization tasks, including the liver, kidneys, and prostate.

B. Statistical Iterative Reconstruction

The reconstructed image $\hat{\mu}$ is solved from projection measurements y by iterative updates using the parallelizable separable quadratic surrogates (SQS) technique [4] to solve the PL optimization problem in (1), which balances data consistency with image regularization.

$$\hat{\mu} = \underset{\mu}{\operatorname{argmax}} L(\mu; y) - \beta R(\mu) \tag{1}$$

The log-likelihood term L maximizes consistency of the forward-projected reconstructed image $\mathbf{A}\mu$ with the measurements while simultaneously accounting for the statistics of the projection data by modeling the Poisson distribution of all measurements. The effective I_0 was estimated from mean and variance measurements of projections in air.

$$L(\mu; y) = \sum_{i \in P} y_i \log\left(I_{0,i} e^{-[\mathbf{A}\mu]_i}\right) - I_{0,i}\, e^{-[\mathbf{A}\mu]_i} \tag{2}$$

The strength of image regularization $R(\mu)$ is controlled by the parameter β and depends on the penalty function $\psi(x)$ of pixel signal differences with first-order neighbors.

$$R(\mu) = \sum_{i \in I} \sum_{j \in N} \psi(\mu_i - \mu_j) \tag{3}$$

While regularization seeks to reduce noise, its behavior depends not only on β, but also the function $\psi(x)$, which we take to be the edge-preserving Huber penalty (PL-H).

$$\psi_H(x) = \begin{cases} \frac{1}{2\delta} x^2, & |x| \le \delta \\ |x| - \frac{\delta}{2}, & |x| > \delta \end{cases} \tag{4}$$

This Huber function is parameterized by a transition width δ, below which a quadratic penalty is applied and above which a linear penalty (similar to total variation, TV [5]) is applied. A smaller value of δ provides a TV-like penalty that favors edges and uniform regions, while a larger δ tends toward a quadratic penalty (PL-Q). The parameter δ carries the same units as the reconstruction (attenuation coefficient, mm^{-1}) and is converted to Hounsfield units (HU) by an approximate factor of 5×10^4.

The optimization was solved by iteratively updating the reconstructed volume using in-house GPU-implemented (GTX680, nVidia) forward- and back-projection operations using the separable footprints projection method for increased accuracy [6]. Reconstruction time for the current (un-optimized) implementation remains time-consuming (~150 min for 500×400×80 voxel reconstructions, run for 300 iterations with ordered subsets to ensure convergence) but will improve with optimized implementation and advances in hardware since computation speed improves approximately linearly with the number of processing units (CUDA cores) by virtue of the highly parallelized implementation.

To handle the large degree of truncation due to the limited field of view, a technique was devised in which a virtual elliptical water cylinder (Fig. 1) of size and location determined from the projection data was estimated by active-set minimization of the mean squared error between the line integrals of the cylinder and the projection data. This elliptic cylinder was then used to initialize PL reconstruction to remove truncation artifacts and improve the reconstructed accuracy within the C-arm FOV. The technique provided superior initialization compared to an FBP reconstruction by giving a better support region outside of the C-arm FOV.

C. Image Quality Assessment

Fair comparison of image quality performance in PL and FBP involves careful "matching" of spatial resolution or noise, such as comparison of contrast-to-noise ratio (CNR) performed at equivalent spatial resolution, since each approach has parameters allowing tradeoff of noise and resolution. At matched and fixed spatial resolution, simple metrics such as the CNR of the low-contrast target provide a basic comparison of low-contrast visibility and a starting point for assessing the potential improvement in image quality at different dose levels. The edge spread function (ESF) across the low-contrast sphere was used as a metric of spatial resolution. An error function (erf) with width parameterized by σ (mm) was fitted to the reconstructed voxel values as a function of distance from the center of the sphere in a narrow conical region, an extension of the 2D case in [7]. The width σ of the erf fit provided a basic measure of resolution length, while its height provided a measure of contrast seen across the boundary. Noise was measured in a small volume of the liver region adjacent to the low-contrast sphere.

The noise-resolution tradeoff with respect to the low-contrast sphere was captured by varying the PL regularization parameter β and the FBP apodization cutoff frequency. Since PL regularization was applied in all directions in a 3D neighborhood, a 2D apodization kernel was applied in the FBP filtration step to provide fair matching of 3D spatial resolution.

III. Results

A. PL variants at low dose

As shown in Fig. 2, reconstructions of the low-dose 60 mAs scan of the phantom with PL-H (δ ranging from 1 to 20 HU), PL-Q, and FBP exhibited a noise-resolution tradeoff that clearly illustrated the noise reduction properties of the Huber

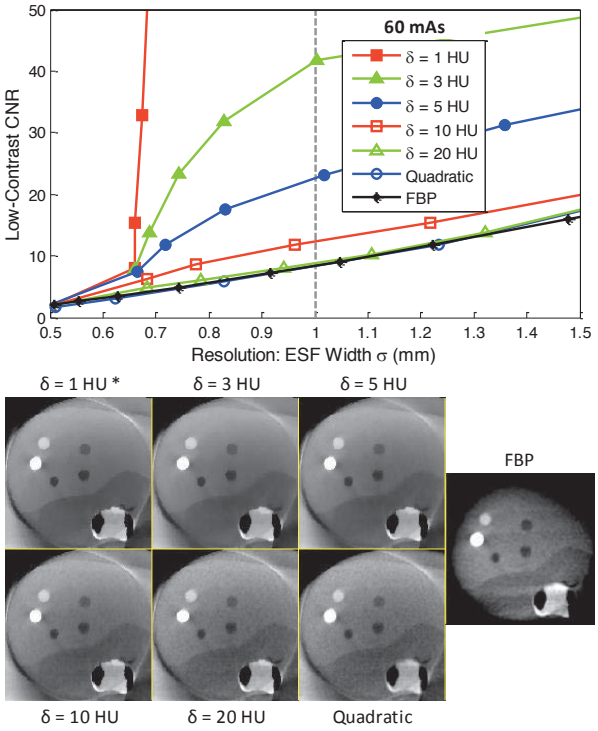


Fig. 2. (Top) Tradeoff between CNR and resolution (ESF) with respect to the low-contrast sphere for PL-H with different choices of δ, PL-Q, and FBP at 60 mAs. (Bottom) Axial slices of the reconstructions for ESF σ = 1 mm.
* δ = 1 HU is shown at higher resolution (σ = 0.7 mm) due to the high degree of edge preservation.

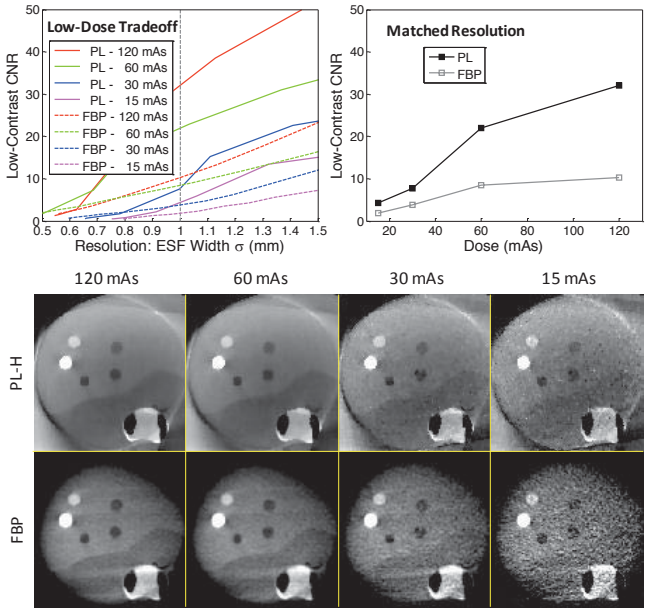


Fig. 3. (Top left) Tradeoff in CNR and resolution for both PL and FBP at different low dose levels. (Top right) The CNR from PL remained 2-3× higher than FBP for different dose levels at fixed ESF σ = 1 mm. (Bottom) Axial slices of PL (δ = 5 HU) and FBP reconstructions at σ = 1 mm.

penalty function (Fig. 2), especially with smaller δ. Selecting δ = 5 HU for soft-tissue tasks represented a reasonable tradeoff between the desired gain in CNR and avoidance of qualitatively patchy texture often seen in TV approaches and was used in subsequent results.

PL-Q and PL-H with δ ≥ 20 HU yielded similar performance to FBP with respect to low-contrast tasks. The edge of high contrast spheres may still be in the linear regime of the Huber penalty function and will be better preserved than with a quadratic penalty or FBP. While PL-Q and FBP have a simple, almost linear relationship between CNR and ESF, the Huber penalty offered a rapid increase in CNR with lower spatial resolution at a resolution threshold that was dependent on the object, δ, and dose (as discussed below).

An example of matched spatial resolution (ESF σ = 1 mm) is shown in the axial slices of the reconstructed volumes in Fig. 2, displaying the reduced noise of the Huber penalty. Because resolution is contrast- and location-dependent, the resolution was only considered matched with respect to the low-contrast sphere, which is indicative of soft-tissue performance.

B. *Low dose limits*

As the dose was reduced by factors of 2× and 4× by lowering the tube current, PL (Huber, δ = 5 HU) continued to show higher CNR than FBP at the same spatial resolution. While CNR also dropped for PL at lower dose, it remained 2.0× higher at the lowest dose (15 mAs), up to 3.1× higher for higher dose (120 mAs) at σ = 1 mm resolution (Fig. 3). PL had a greater relative advantage at higher dose, where the low-contrast target was better defined, allowing for more aggressive noise reduction. At lower dose, the benefit from PL could only be realized at lower resolution, when the image was more heavily smoothed and regularization could suppress background noise. The greater relative advantage of PL over FBP was therefore maintained in the regime above 30 mAs in this scenario. While CNR is a simplistic image quality metric, the results nonetheless suggest PL offers the potential to image at lower dose due to the improved image quality. For example, PL images at 60 mAs yield greater CNR than FBP at 120 mAs. Such dose reduction must of course be considered alongside rigorous assessment of the clinical acceptability of the images – noise in the 15 mAs scan may be too high to be clinically acceptable, even with PL.

C. *Sparse projections*

Fig. 4 compares PL imaging performance and reconstructions from the fully sampled 60 mAs acquisition with a sparse data set, where half the projections were taken from the 120 mAs acquisition (net 60 mAs), yielding the same dose. The undersampled data generally had lower CNR than the fully sampled data of the same net dose, though it remained well above that of the fully sampled data at half the dose (30 mAs). At lower resolution (σ > 1.1 mm), the sparse projections still provided sufficient sampling for the PL-H reconstruction to yield comparable or slightly higher CNR due to the reduced effect of electronic noise. Axial images (Fig. 4) demonstrated that while the Huber penalty better suppressed undersampling streak artifacts compared to the quadratic penalty, concentrating the dose into fewer projections yielded lower overall performance than lowering the dose per projection with a full data set.

D. *Cadaver images*

Applying the same reconstruction parameters yielding σ = 1 mm resolution for low-contrast structures in the phantom to reconstructions of abdominal scans of a cadaver demonstrated the promise of PL in soft-tissue imaging in a real anatomical context. While the appropriate parameters generally depend on

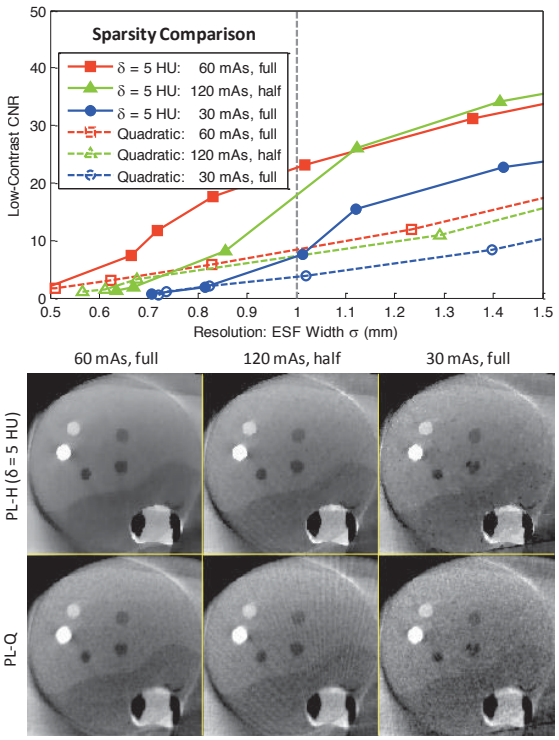


Fig. 4. (Top) Tradeoff in CNR and resolution for PL-H and PL-Q, comparing tube current reduction (60 mAs, fully sampled) with undersampled acquisitions (half of the 120 mAs projections). (Bottom) Axial slices of the reconstructions demonstrate that PL-H better reduced the view aliasing artifacts from undersampled projections.

the object size, dose, and contrast associated with the imaging task, the cadaver was approximately the same size as the phantom, and the imaging task remained that of low-contrast visualization. A reconstruction for a normal dose scan at 230 mAs provided reference images, and PL reconstructions at 60 mAs and 30 mAs (fully sampled) compared favorably with FBP in delineating the boundary of the kidney and other soft-tissues, as well as the fine, high contrast structures below and in the spine. The PL images demonstrated robustness to noise and artifacts that obscured soft-tissue visibility in FBP. Additionally, because the FBP images were resolution-matched for soft-tissues, the high contrast structures were also reconstructed at the same spatial resolution and did not enjoy the edge-preservation provided by PL-H.

The sparse data set, though having the same net dose as the fully sampled, 60 mAs acquisition and offering some suppression of undersampling artifacts present in FBP, remained less favorable than the fully sampled data sets. The results therefore suggest that reducing dose via fully sampled, lower mAs acquisitions offers greater potential for dose reduction within this low-dose range.

IV. Discussion and Conclusion

Application of advanced 3D image reconstruction methods such as PL holds considerable promise for low-dose, soft-tissue imaging in C-arm CBCT. In the imaging scenarios examined and with the current level of electronic noise exhibited by flat-panel detectors, low-dose acquisition exhibited better image quality through tube current reduction, rather than sparse sampling. A regime where sparse sampling is expected to have greater advantage is when additive noise sources is the dominant source of noise, and concentrating

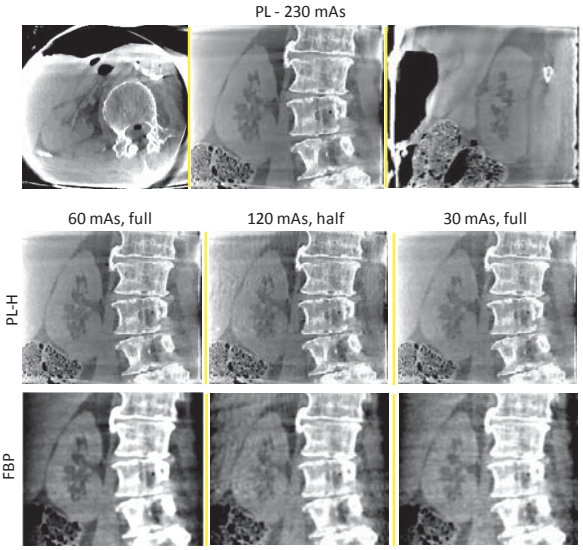


Fig. 5. (Top) Axial, coronal, and sagittal slices of the cadaver's right kidney at standard dose (230 mAs). (Middle) Coronal slices of PL-H (δ=5 HU) at 60 mAs, fully sampled; half of the projections from 120 mAs (net: 60 mAs); 30 mAs, fully sampled. (Bottom) FBP with matched low-contrast resolution (σ=1 mm) as determined from the phantom.

dose into fewer projections allows for higher SNR. Measurements on the current system showed electronic noise $\sigma_e^2 = 5.8$ analog-to-digital units (ADU), a small contribution compared to the quantum noise (typically, $\sigma_q^2 > 25$ ADU) under the conditions in this study. As more advanced modeling of the imaging chain enables PL reconstructions of ultra-low-dose protocols, sparse sampling may become more advantageous.

The reduction of image noise provided by PL over FBP (at matched spatial resolution) in low-dose scans was quantitatively demonstrated in an anthropomorphic phantom, and the resulting improvement was validated in cadaver studies. The results demonstrated superior soft-tissue visualization with PL reconstruction even at lower dose than FBP, including improved delineation of the kidney and surrounding soft-tissues. Ongoing research will investigate the low-dose and low-contrast limits made possible by more advanced PL reconstruction that incorporates improvements such as scatter artifact correction, an electronic noise model, and enhancement of reconstruction speed to further address factors that conventionally limit CBCT soft-tissue imaging performance.

References

[1] J. Siewerdsen, *et al.*, "Volume CT with a flat-panel detector on a mobile, isocentric C-arm: pre-clinical investigation in guidance of minimally invasive surgery," *Med. Phys.*, vol. 32, p. 241, 2005.

[2] G. H. Chen, *et al.*, "Design and development of C-arm based cone-beam CT for image-guided interventions," *Proc. SPIE*, 2006, p. 614210.

[3] Y. Sagara, *et al.*, "Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients," *Am. J. Roentgen.*, vol. 195, p. 713, 2010.

[4] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, p. 2835, 1999.

[5] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, p. 4777, 2008.

[6] Y. Long, *et al.*, "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.*, vol. 29, p. 1839, 2010.

[7] J. D. Evans, *et al.*, "Noise-resolution tradeoffs in x-ray CT imaging: A comparison of penalized alternating minimization and filtered backprojection algorithms," *Med. Phys.*, vol. 38, p. 1444, 2011.

Space-Variant Channelized Preconditioner Design for 3D Iterative CT Reconstruction

Lin Fu, Zhou Yu, Jean-Baptiste Thibault, Bruno De Man, Madison G. McGaffin, and Jeffrey A. Fessler

***Abstract*— Preconditioners, especially diagonal ones, have been key ingredients in several state-of-the-art optimization algorithms for transmission and emission tomographic image reconstruction. But it remains challenging to design robust, non-diagonal preconditioners that can account for various space-variant factors such as fan-beam/helical sampling geometry, non-uniform statistical weights, and object-dependent regularization. In this study, we propose a channelized preconditioner design that decomposes a preconditioner into multiple channels that represent different frequency sub-bands and/or orientations. Each channel is associated with a spatial weighting map to modulate its gain at different spatial locations. The multi-channel design has the potential to provide more degrees of freedom in controlling the localized spectral response of the preconditioner without incurring excessive computational overhead. Initial application to maximum *a posteriori* probability image reconstruction from helical x-ray CT data is presented here.**

***Index Terms*— computed tomography, iterative image reconstruction, preconditioner.**

I. Introduction

Model-based iterative reconstruction (MBIR) techniques for x-ray computed tomography (CT) have been developed over a decade ago [1], [2], but have only been recently introduced commercially on multi-slice clinical CT scanners. Based on the principles of maximum *a posteriori* probability (MAP) estimation, the model-based approach improves multiple aspects of image quality, and has demonstrated potential dose savings in recent clinical trials compared to the conventional filtered backprojection (FBP) method and other state-of-the-art CT reconstruction methods [3], [4].

Due to the complexity of various geometrical, physical, and statistical models being employed by MBIR, and the large size of data acquired by today's multi-slice CT scanners, the computational cost of MBIR remains a major impediment to its widespread use in clinical environments. It is a topic of growing interest to develop accelerated MBIR algorithms. This study will focus on gradient-based simultaneous-update optimization algorithms, which have relatively high level of parallelism and could potentially take full advantage of many-core computing devices.

The project described was supported by Grant Number 1-R01-HL-098686 from NIH and its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

L. Fu and B. De Man are with GE Global Research Center, Niskayuna NY 12309. Z. Yu and J.-B. Thibault are with GE Healthcare Technologies, Waukesha, WI 53188. M. G. McGaffin and J. A. Fessler are with the Electrical Engineering and Computer Science department, University of Michigan, Ann Arbor, MI 48109.

Standard gradient-based iterations usually converge slowly for large-scale ill-conditioned problems. Effective preconditioning techniques are essential for their acceleration and practical success. Various forms of preconditioners have been studied in the context of iterative tomographic reconstruction. Diagonal scaling matrix is the simplest form of preconditioner. Several widely used iterative algorithms in emission and transmission reconstruction can be viewed as diagonally-preconditioned gradient descent algorithms (for instance EM [5], [6], SQS [7], and ML-TR [8]). Diagonal preconditioners are also commonly combined with other optimization algorithms such as conjugate-gradient to achieve more significant acceleration [9–11]. Despite their effectiveness and robustness, diagonal preconditioners are considered relatively conservative approximations to the inverse of the Hessian matrix and can only provide suboptimal acceleration.

Non-diagonal, Fourier preconditioners have the potential to address the off-diagonal structure of the Hessian. Such preconditioners are also attractive because of their close connection to the ramp-filter used in FBP reconstruction. These preconditioners can bring dramatic acceleration for space-invariant problems [12], but they are less effective for space-variant reconstruction due to factors such as irregular geometric sampling, non-uniform statistical noise modeling, and location-dependent image priors.

To improve convergence rates in space-variant reconstruction, Booth and Fessler proposed a preconditioning technique based on the product of a Fourier kernel and a particular diagonal matrix [11]. Such combined preconditioner yields significantly faster convergence than either Fourier or diagonal preconditioning alone. In a subsequent work, the Fourier component was further generalized by interpolation among multiple FFTs to provide more effective handling of space-variant regularization strength [13]. More recently, operator splitting methods has been proposed for preconditioners to better address shift-variant problems [14], [15].

Some promising results have recently been reported in applying preconditioning techniques to accelerating CT MBIR [16]. It will be of great interest to develop more effective and efficient preconditioners for multi-slice medical CT systems where the geometric sampling can be incomplete or truncated, and the statistical noise is highly anisotropic. To properly account for the space-variant effects, we would need a sufficient degree of freedom in controlling the local frequency response of the preconditioner without incurring excessive computational overhead. In this study, we propose a channelized preconditioner design, in which the preconditioner is decomposed into different channels representing different frequency sub-bands and/or orientations. A spatial

weighting map is applied to each channel to modulate its gain at different locations. While single channel may be restrictive in modeling space-variant frequency response, the combination of multiple channels may provide a sufficient degree of freedom in controlling the local frequency response without incurring excessive computation overhead.

II. Theory

A. MAP cost function

One approach to statistical image reconstruction in x-ray CT uses a MAP cost function in the form of

$$\Phi(\boldsymbol{x}) = -L(\mathbf{A}\boldsymbol{x}; \boldsymbol{y}) + U(\boldsymbol{x}),$$

where $\boldsymbol{x} = \{x_1, \dots, x_N\}$ denotes the vector of unknown 3D image space; $\boldsymbol{y} = \{y_1, \dots, y_M\}$ is the vector of sinogram measurements; $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ denotes the system matrix; $-L(\cdot,\cdot)$ is the negative log likelihood term that penalizes the inconsistency between the estimated projection data and the physical measurements; $U(x)$ is the regularization function that penalizes the noise in the image.

In this study, we use Gaussian log likelihood function with the noise covariance matrix $\mathbf{W}^{-1}$ and $L(\mathbf{A}\boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{x})^T\mathbf{W}(\boldsymbol{y} - \mathbf{A}\boldsymbol{x})$. The regularization function $U(\boldsymbol{x})$ is expressed by a Markov Random Field (MRF) in the form of $U(\boldsymbol{x}) = \sum_{k \in \mathcal{N}_j, k>j} \omega_{jk}\rho(x_j - x_k)$, with $\mathcal{N}_j$ denoting the collection of the neighboring pixels for location j, ω_{jk} representing the penalty strength between pixel j and k, and $\rho(\cdot)$ being a prior potential function. We use the q-GGMRF prior with $\rho(\Delta) = \frac{|\Delta|^p}{1+\left|\frac{\Delta}{c}\right|^{p-q}}$ and $1 \le q \le p \le 2$ to ensure convexity [2].

B. Hessian matrix and local spectral analysis

The Hessian matrix for the MAP cost function is

$$\mathbf{H}(\boldsymbol{x}) \triangleq \nabla^2\Phi(\boldsymbol{x}) = \mathbf{A}^T\mathbf{W}\mathbf{A} + \nabla^2 U(\boldsymbol{x}).$$

We would like the preconditioner $\mathbf{M}$ to be an effective approximation to the inverse of the Hessian matrix, so that $\mathbf{MH} \approx \mathbf{I}$, or the condition number of $\mathbf{MH}$ be minimized. This direct matrix optimization problem seems not tractable so approximations have to be used.

Based on the concept of local shift invariance [17], [18], it is generally assumed that the Hessian matrix is locally block-Toeplitz, so that it can be approximately diagonalized by Fourier transforms. The local spectral representation of $\mathbf{H}(\boldsymbol{x})$ at the location of the jth pixel is

$$\begin{aligned}\boldsymbol{h}^j(\boldsymbol{x}) &\triangleq \mathbf{H}(\boldsymbol{x})\boldsymbol{e}^j \\ &= \mathbf{Q}^T\mathrm{diag}\{\boldsymbol{\lambda}^j + \boldsymbol{\mu}^j(\boldsymbol{x})\}\mathbf{Q}\boldsymbol{e}^j,\end{aligned}$$

where $\boldsymbol{h}^j(\boldsymbol{x})$ is the jth column of $\mathbf{H}(\boldsymbol{x})$; $\boldsymbol{e}^j$ is the jth unit vector; $\mathbf{Q}$ represents a discrete Fourier transform; $\boldsymbol{\lambda}^j$ and $\boldsymbol{\mu}^j$ are the Fourier transforms of the jth column of $\mathbf{A}^T\mathbf{W}\mathbf{A}$ and $\nabla^2 U(\boldsymbol{x})$, respectively.

$$\begin{aligned}\boldsymbol{\lambda}^j &= \mathrm{diag}\{\mathbf{Q}\boldsymbol{e}^j\}^{-1}\mathbf{Q}[\mathbf{A}^T\mathbf{W}\mathbf{A}\boldsymbol{e}^j], \\ \boldsymbol{\mu}^j(\boldsymbol{x}) &= \mathrm{diag}\{\mathbf{Q}\boldsymbol{e}^j\}^{-1}\mathbf{Q}[\nabla^2 U(\boldsymbol{x})\boldsymbol{e}^j].\end{aligned}$$

The term $\mathrm{diag}\{\mathbf{Q}\boldsymbol{e}^j\}^{-1}$ adds the appropriate complex exponentials.

C. Ramp-filter based preconditioner

In a previous study [16], we designed a ramp-filter based preconditioner based on the combined diagonal/circulant formulation proposed by Fessler and Booth [13]. In this study, we implement a similar but improved design, and will compare its performance to the channelized preconditioner to be introduced in the next few sections. The ramp-based preconditioner in this study takes the form of

$$\mathbf{M}_{\mathrm{ramp}} = \mathrm{diag}\left(\Lambda_j\right)^{-1}\mathbf{K}\mathrm{diag}\left(\Lambda_j\right)^{-1},$$

where $\mathbf{K}$ is a space-invariant isotropic Fourier kernel, and Λ_j is a particular spatial weighting factor that makes the preconditioner space-variant [13]. The filter kernel $\mathbf{K}$ is designed based on a continuous-space approximation to the matrix spectra $\boldsymbol{\lambda}^j$ and $\boldsymbol{\mu}^j$ [19]. The frequency response of the filter resembles an apodized ramp-filter

$$K(e^{j2\pi f}) = \left\{\frac{1}{|f| + \delta_0} + \left(\frac{\kappa_0}{\Lambda_0}\right)^2 2\sin^2(\pi f)\right\}^{-1},$$

where $-0.5 < f < 0.5$ is normalized digital frequency, and Λ_0, δ_0, and κ_0 are parameters to adjust the shape of the frequency response. The filter kernel can be adjusted to match the Hessian only at a single location, thus it is restrictive in modeling highly space-variant system response.

D. Channelized preconditioner

To provide more degrees of freedom in approximating the space-variant factors in the Hessian, we propose a channelized preconditioner design, in which the preconditioner consists of K predetermined frequency channels:

$$\mathbf{M}(\boldsymbol{x}) = \sum_{k=1}^{K} \mathrm{diag}\{\boldsymbol{t}_k(\boldsymbol{x})\}\mathbf{M}_k\mathrm{diag}\{\boldsymbol{t}_k(\boldsymbol{x})\},$$

where each channel represents a frequency sub-band and/or spatial orientation. $\mathbf{M}_k$ is a positive-definite filter that defines the frequency response of the kth channel, and $\boldsymbol{t}_k$ is a spatial weighting map that modulates the gain of the kth channel at different locations. By splitting the preconditioner into different channels, we could control the gain of each channel independently, which gives the potential to incorporate more space-variant effects. The ramp-based preconditioner introduced earlier can be viewed as a special case where a single high-pass channel is used.

E. Channel design

To design the channelized preconditioner, we first determine the number of channels and the frequency response of each channel. The channels are like basis functions. More channels will improve the frequency resolution but also increase design complexity and computational overhead. In this initial study, making no attempt to optimize the channel

design in general, we explore the possibility of using only three channels ($K = 3$). These channels could be implemented in either frequency domain or space domain. Here we propose image-space kernels with very small footprint (3x3x3), which has less computational overhead than implementing these kernels in Fourier space.

$$\mathrm{Ker}\mathbf{M}_1 = \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 2 & 12 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\}$$

$$\mathrm{Ker}\mathbf{M}_2 = \left\{ \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix}, \begin{bmatrix} -2 & -4 & -2 \\ -4 & 24 & -4 \\ -2 & -4 & -2 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix} \right\}$$

$$\mathrm{Ker}\mathbf{M}_3 = \left\{ \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} -2 & -4 & -2 \\ -4 & 24 & -4 \\ -2 & -4 & -2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix} \right\}$$

The first channel represents low frequency in plane (x-y). The second one represents high frequency in plane (x-y) and low frequency across plane (z). And the third one represents high frequency both in plane (x-y) and across planes (z). These channels are approximately isotropic, i.e., no angular preference. However it is possible to design channels with different spatial orientations to account for anisotropic effects in the Hessian. This will be a potential topic for further investigation. As will be shown in the initial results here, even the relatively simple three-channel design can provide significant improvement over the conventional diagonal or combined diagonal/circulant preconditioners.

F. Spatial weighting design

Now we will design the spatial weighting $\boldsymbol{t}_k$ associated with each channel. We would like to design $\boldsymbol{t}_k$ so that

$$\mathbf{M}(\boldsymbol{x})\mathbf{H}(\boldsymbol{x}) \approx \mathbf{I}.$$

In local spectral representation, this condition becomes

$$\left[\mathbf{Q}^T \left(\sum_{k=1}^{K} t_{kj}^2 \mathrm{diag}\{\boldsymbol{\nu}_k\}\right) \mathbf{Q}\right] [\mathbf{Q}^T \mathrm{diag}\{\boldsymbol{\lambda}^j + \boldsymbol{\mu}^j(\boldsymbol{x})\}\mathbf{Q}] \approx \mathbf{I},$$

where $\boldsymbol{\nu}_k$ denote the Fourier transform of $\mathbf{M}_k$, i.e, $\mathbf{M}_k = \mathbf{Q}^T \mathrm{diag}\{\boldsymbol{\nu}_k\}\mathbf{Q}$, and t_{kj} is the jth element of the $\boldsymbol{t}_k$, representing the spatial weighting for the kth channel at location j. After simplification, we obtain a system of linear equations that $\boldsymbol{t}_k$ should satisfy:

$$\sum_{k=1}^{K} \boldsymbol{\nu}_k \odot (\boldsymbol{\lambda}^j + \boldsymbol{\mu}^j(\boldsymbol{x})) t_{kj}^2 \approx \mathbf{1},$$

where " $\odot$ " denote element-wise multiplication. This condition means we would like the preconditioned problem to have a "flat" spectrum. The equation is over-determined and does not have a strict solution. Although various generalized inverse or optimization based methods may be used to solve such an over-determined problem, here we propose to find an approximate solution that only satisfies the equations at K frequency samples, $\boldsymbol{f}_1 \ \ldots \ \boldsymbol{f}_K$. In matrix form this yields

$$\begin{bmatrix} \nu_{11} & \cdots & \nu_{1K} \\ \vdots & \ddots & \vdots \\ \nu_{K1} & \cdots & \nu_{KK} \end{bmatrix} \begin{bmatrix} (\lambda_1^j + \mu_1^j) t_{1j}^2 \\ \vdots \\ (\lambda_K^j + \mu_K^j) t_{Kj}^2 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

where $\nu_{ik} \triangleq \nu_i(\boldsymbol{f}_k)$, $\lambda_k^j \triangleq \lambda^j(\boldsymbol{f}_k)$, $\mu_k^j \triangleq \mu^j(\boldsymbol{f}_K, \boldsymbol{x})$. Assuming the frequency channels are linearly independent, the solution is obtained by inverting the K-by-K matrix formed by $\{\nu_{ik}\}$

$$\begin{bmatrix} t_{1j}^2 \\ \vdots \\ t_{Kj}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1^j + \mu_1^j & & \\ & \ddots & \\ & & \lambda_K^j + \mu_K^j \end{bmatrix}^{-1} \begin{bmatrix} \nu_{11} & \cdots & \nu_{1K} \\ \vdots & \ddots & \vdots \\ \nu_{K1} & \cdots & \nu_{KK} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

If the channels have good frequency separation, we may further assume the K-by-K matrix is approximately diagonal, which yields a simple approximate solution:

$$t_{jk}^2 = \frac{1}{(\lambda_k^j + \mu_k^j)\nu_{kk}}.$$

Computing $\boldsymbol{t}_k$s using the above expression still requires $\lambda_k^j + \mu_k^j$ be obtained first. Efficient methods to compute λ_k^j have been proposed, where computation may be performed only at sparsely sampled locations and then be interpolated to others [20][21]. λ_k^j and μ_k^j may also be obtained using continuous-space approximations.

The spectral coefficients $\boldsymbol{\lambda}^j$ are not only space-variant but also anisotropic due to the high dynamic range of the statistical weights $\mathbf{W}$. Since we are only using approximately isotropic channels in this initial study, we design the preconditioner by matching the frequency response in the angular direction associated with the strongest statistical weights. This is a more conservative choice compared with matching the averaged frequency response over all angular directions, and helps make the preconditioner more robust. We may also apply empirical adjustment to gain of the spatial-frequency channels. For example, we may reduce the gain of high frequency channels in the regions where the sampling is incomplete and the Hessian is highly space-variant.

III. APPLICATION TO HELICAL CT DATA

We tested the algorithms using a chest scan acquired on a 64-slice GE HD750 CT scanner at 120 kV with a helical pitch of one. The reconstruction field-of-view was 70 cm in diameter and 8.5 cm in Z direction, with image matrix size of 600 x 600 x 136. We compared the convergence rates of four numerical algorithms: standard conjugate-gradient (CG), CG with separable quadratic surrogate (SQS) preconditioner [7], CG with ramp-based preconditioner, and CG with the proposed 3-channel preconditioner. All algorithms are initialized with standard FBP reconstructions and run with 10 iterations. An approximately fully converged reference reconstruction is generated by 20 iterations of non-homogenous iterative coordinate descent (ICD) [2], [22].

Sample images reconstructed from different methods are shown in Figure 1. At only 10 iterations, the proposed channelized preconditioner generates images that have similar visual quality as the reference MBIR solution, while other reconstruction methods show less satisfactory image

quality. Figure 2 shows the image-domain L2 distance to the reference solution as a function of iteration number. The channelized preconditioner clearly achieves the fastest convergence rate.

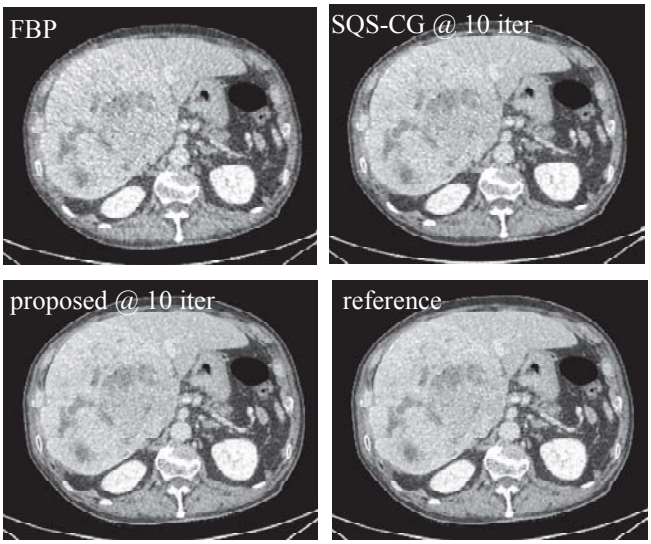


Figure 1. Sample reconstructions with different methods. All reconstructions are initialized with the FBP image. Display window = [-200 200] HU.

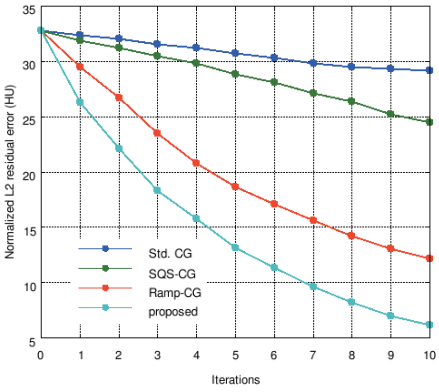


Figure 2. Convergence curves.

IV. Summary and Discussion

We have presented a channelized preconditioner design for CT MBIR problems. In the proposed design, the channels represent different spatial frequency sub-bands, and the gain for each channel is space-variant and independently modulated by a spatial weighting map. The introduction of the multiple channels has the potential to provide more degrees of freedom in approximating space-variant factors in the Hessian matrix. Compared to the channelized design, the conventional diagonal preconditioners can be viewed as a single all-pass channel, while the ramp-based preconditioner can be viewed as a single high-pass channel, both being special cases of the proposed channelized design.

Unlike previous methods based on FFTs which aim to model the frequency response accurately, we recognize the tradeoff among frequency resolution, space-variance, and computational cost, and used very small 3x3x3 preconditioner kernels which bring little computational overhead to MBIR. The new algorithm is tested with helical CT data and effective acceleration compared to other conventional types of diagonal and Fourier preconditioners is illustrated.

V. References

[1] B. De Man and J. A. Fessler, "Statistical iterative reconstruction for X-ray computed tomography," in *Biomedical Mathematics*, G. Wang, Y. Censor, and J. Ming, Eds. Springer, 2009.

[2] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[3] P. J. Pickhardt, M. G. Lubner, D. H. Kim, J. Tang, J. A. Ruma, A. M. del Rio, and G.-H. Chen, "Abdominal CT with model-based iterative reconstruction (MBIR): initial results of a prospective Trial comparing ultralow-dose with standard-dose imaging," *American Journal of Roentgenology*, vol. 199, p. to appear, 2012.

[4] A. Neroladaki, D. Botsikas, S. Boudabbous, C. D. Becker, and X. Montet, "Computed tomography of the chest with model-based iterative reconstruction using a radiation exposure similar to chest X-ray examination: preliminary observations," *European Radiology*, p. ahead of print, available online, 2012.

[5] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, pp. 306–316, 1984.

[6] L. Kaufman, "Implementing and accelerating the em algorithm for positron emission tomography.," *IEEE Transactions on Medical Imaging*, vol. 6, no. 1, pp. 37–51, 1987.

[7] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–51, 1999.

[8] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 999–1008, 2001.

[9] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 687–701, 1994.

[10] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small animal scanner," *Phys. Med. Biol.*, vol. 43, pp. 1001–13, 1998.

[11] S. D. Booth and J. A. Fessler, *Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography*, vol. 2. 1995, pp. 441–444.

[12] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions.," *IEEE Trans. Med. Imag.*, vol. 12, no. 1, pp. 78–83, 1993.

[13] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Imag. Proc.*, vol. 8, no. 5, pp. 688–99, 1999.

[14] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–88, 2012.

[15] M. G. McGaffin, S. Ramani, and J. A. Fessler, "Reduced memory augmented Lagrangian algorithm for 3D iterative x-ray CT image reconstruction," in *Proc. SPIE 8313*, 2012, p. 831327.

[16] L. Fu, B. De Man, K. Zeng, T. M. Benson, Z. Yu, G. Cao, and J.-B. Thibault, "A Preliminary Investigation of 3D Preconditioned Conjugate Gradient Reconstruction for Cone-Beam CT," in *Proc. SPIE. 8313*, 2012, p. 83133O.

[17] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Imag. Proc.*, vol. 5, pp. 1346–1358, 1996.

[18] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, 2000.

[19] J. A. Fessler, "Analytical approach to regularization design for isotropic spatial resolution," in *IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2003, pp. 2022–6.

[20] J. W. Stayman and J. A. Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems.," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 269–284, 2004.

[21] L. Yang, J. Zhou, and J. Qi, "Penalized maximum-likelihood image reconstruction for 3D breast lesion detection," in *IEEE Nuclear Science Symposium Conference*, 2012.

[22] Z. Yu, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization," *IEEE Trans. Imag. Proc.*, vol. 20, no. 1, pp. 161–175, 2011.

Fully 6D image reconstruction for myocardial perfusion imaging of tracer dynamics, cardiac and respiratory motion

Uttam Shrestha, Fares Alhassen, Rostyslav Buchko, Robert G. Gould, Youngho Seo, *Senior Member, IEEE*, Elias H. Botvinick, and Grant T. Gullberg, *Fellow, IEEE*

***Abstract* – Dynamic data acquisition and spatiotemporal reconstruction of the myocardial perfusion imaging (MPI) using single photon emission computed tomography (SPECT) have been drawing a significant interest for a decade. However, time-varying nature of the radiopharmaceuticals as well as unavoidable involuntary motion of the heart due to quasiperiodic beating and the effects of respiration and diaphragm motion usually mar the quality of the image in this modality. In this work, we developed a technique to reconstruct an accurate spatiotemporal distribution of the radiopharmaceutical in the myocardium by using the tensor product of different sets of basis functions that approximately describe the internal motion of the heart. The temporal basis functions were chosen to reflect the optimal tracer dynamics while the motion of the heart were described by a discrete set of cardiac and respiratory states. The voxelized three-dimensional (3D) physical space, the temporal variation of the tracer, and the gated cardiac and respiratory motion make the calculation fully six-dimensional (6D).**

Index Terms* - Myocardial perfusion, dynamic-SPECT, motion correction, fully 6D.

I. Introduction

Noninvasive myocardial perfusion imaging (MPI) has been a useful tool for both diagnostic and prognostic assessment of patients suffering from coronary syndromes including ischemia and myocardial dysfunction. There are a number of imaging modalities such as x-ray computer tomography (CT), positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) that have been used to assess the critical anatomical and functional impairments of the heart. In particular, emission tomography using dynamic PET and dynamic SPECT has been and continues to play a leading role in the clinical noninvasive diagnosis of myocardial ischemia [1]. However, despite widespread applications, images are often degraded by the involuntary motion of the heart leading to blurring, and sometimes to erroneous diagnosis. Cardiac gating, which is generally considered a practical method to assess the myocardial motion [2], lacks the simultaneous assessment of the temporal variation of the tracer. The main goal of this study is to simultaneously model the motion of the human heart (that includes both cardiac beating and motion due to respiration), and the distribution of the radiotracer for the dynamical MPI using SPECT.

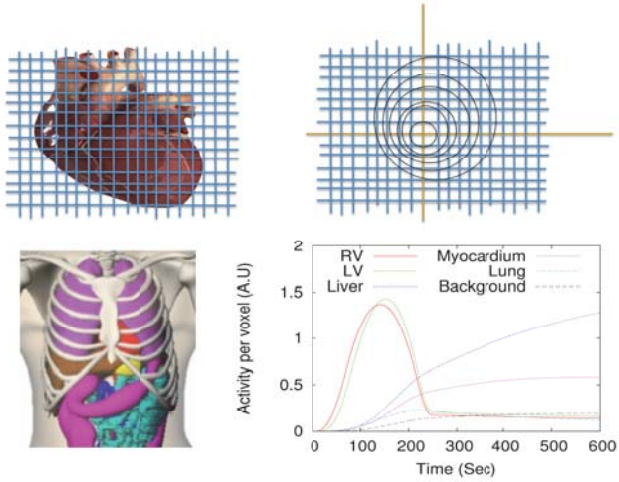


Fig. 1. Portrayal of cardiac states (top row), and a phantom torso [14] showing organs and corresponding simulated time-activity curves (bottom row).

In this work, we formulate the reconstruction algorithm for the projection of dynamic cardiac SPECT data using the tensor product of the spatiotemporal basis functions and the basis functions of the subspaces of the cardiac and respiratory phases. Time-varying activity within a volume, in practice, can be modeled by selecting a set of temporal basis functions that are capable of representing approximate local time variations, and have desired smoothness properties. The functions such as splines, factor analysis of dynamic structures (FADS) [3], or other possible spectral decompositions [4] [5] are a few examples. Similarly, the spatially nonuniform activity concentration within a particular volume can be modeled by selecting an appropriate set of spatial basis functions defined within the volume. These could also be splines [6], point clouds of tetrahedral elements [7], blobs [8], various types of polynomial expansions or indicator functions corresponding to voxels. Furthermore, the basis function for the motion can be anything that optimally characterizes the movement of the heart. For instance, the quasiperiodic beating

Manuscript received January 28, 2013. This work was supported in part by
U. Shrestha, F. Alhassen, R. Gould., Y. Seo, and E. Botvinick are with the Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, California, USA (uttam.shrestha@ucsf.edu).

G. Gullberg., and R. Buchko are with the Lawrence Berkeley National Laboratory, Berkeley, California, USA (gtgullberg@lbl.gov).

of the heart can be characterized by some periodic functions that reflect the different cardiac phases, while the motion due to respiration can be some linear translations. The number of basis functions and their spatial and temporal extents can also be varied so that they can optimally model the spatial and temporal content of the data with the fewest number of basis functions.

A smooth time-varying concentration of the tracer activity in a given volume and the deformation of the heart were simultaneously solved using the maximum likelihood expectation maximization (ML-EM) algorithm [9]. The ML-EM algorithm was successfully implemented in the past for modeling four-dimensional (4D) SPECT acquisition [10] as well as for five-dimensional (5D) PET [11] and SPECT acquisitions [12]. Recently, we have reported on a fully 5D reconstruction method using a continuous rotating SPECT system for dynamic MPI [13]. In this study, we extended our previous work to incorporate the effect of respiratory motion. Attenuation correction, a major degrading factor, has been explicitly employed in the calculation through the system matrix while scattering was neglected. Our approach is novel in its own right, and may give a handle to study the internal motion of any organ in a multidimensional space.

II. 6D Model

In dynamic SPECT, a photon emitting radionuclide, e.g., technetium tagged to sestamibi or tetrofosmin, is infused into patient's vein and the emitted photons are detected and recorded continuously by the detector camera as it rotates. The detected photons so called 'projections' provide the information of the kinetics and dynamics of the tracer as well as the motion of the heart.

The activity distribution of a radionuclide in the image space is represented by a function $A(x,t,\tau(t),\zeta(t))$, where x is the spatial coordinate, t is the time coordinate, $\tau(t)$ is the cardiac phase coordinate, and $\zeta(t)$ is the cardiac displacement due to respiratory motion. We model the activity distribution as a tensor product of the spatiotemporal basis functions and the basis functions of the cardiac and respiratory phases:

$$A(x,t,\tau(t),\zeta(t)) = \sum_{m,n,q,r} a_{mnqr} S^m(x) V^n(t) W^q(\tau(t)) R^r(\zeta(t)), \quad (1)$$

where $S^m(x)$, $m = 1,.....,M$, are spatial and $V^n(t)$, $n = 1,....., N$, are temporal basis functions, while $W^q(\tau(t))$, $q = 1,.....,Q$, and $R^r(\zeta(t))$, $r = 1,.....,L$, are basis functions corresponding to the cardiac and respiratory phases, respectively. The expansion coefficients a_{mnqr} give weights for each basis function. Although our formulation is more general, we provide a particular representation of the activity concentration in parametric space to include smooth temporal changes within the volumes when the volume itself is moving in time, and so to provide 6D representations of the dynamic reconstructions.

The projections of the activity at any particular instant depend on the angular position of the detector. The detector is also pixelized so that for an arbitrary i^{th} pixel, accumulation of the photons in a given small time Δt_k at a time point t_k is given by,

$$p_i(t_k,\tau(t_k),\zeta(t_k)) = \int_{t_k-\Delta t_k}^{t_k} \int_{\chi} F[x,d_i(t)] A(x,t,\tau(t),\varsigma(t)) dx dt \quad , \quad (2)$$

where the spatiotemporal distribution of the activity is integrated out along the line in the image space χ. The weighting function $F[x,d_i(t)]$ maps the activity from a position x in the image space into the projection at the detector position d_i. Expanding the activity in terms of the basis functions, the projections can be modeled as

$$p_i(t_k,\tau(t_k),\zeta(t_k)) = \int_{t_k-\Delta t_k}^{t_k} \sum_{r=1}^{R}\sum_{q=1}^{Q}\sum_{n=1}^{N}\sum_{m=1}^{M} a_{mnqr} B^{mnqr}(d_i(t),t,\tau(t),\zeta(t)) dt$$

with

$$B^{mnqr}(d_i(t),t,\tau(t),\zeta(t)) = U^m(d_i(t),t) V^n(t) W^q(\tau(t)) R^r(\zeta(t))$$
$$U^m(d_i(t),t) = \int_{\chi} F[x,d_i(t)] S^m(x) dx . \quad (3)$$

Generally the projection data can be recorded as individual events of radioactive emissions (list mode) or the accumulation of events in a detector bin $d_i(t_k)$ over the acquisition time interval divided by the interval so that the value at all time points are in terms of activity concentration in units of counts per unit time. The projection at time t_k in projection bin I is the activity acquired during the time interval Δt_k:

$$p_i(t_k,\tau_g(t_k),\xi_l(t_k)) = \int_{t_k-\Delta t_k}^{t_k} \sum_{r=1}^{R}\sum_{q=1}^{Q}\sum_{n=1}^{N}\sum_{m=1}^{M} a_{mnqr} B^{mnqr}(d_i(t),t,\tau_g(t),\xi_l(t)) dt$$

where τ_g is the time of the cardiac phase and ξ_l, is the time of the respiratory phase in the interval $[t_k - \Delta t_k, t_k]$ that the projection samples $p(t_k,\tau_g,\xi_l)$, $i = 1,..., I$ are acquired.

We assume there are a sufficient number of cardiac and respiratory cycles during each time interval Δt_k. Since the respiratory period is larger than the cardiac period, for each time interval, we will assume that $L \geq G \geq 1$, where L is the number of respiratory gates $l = 1,..., L$ and G is the number of cardiac gates $g = 1,..., G$ during the respective respiratory and cardiac cycles. The total number of time samples $k = 1,..., K$ is K.

The expression in (3) can be represented in matrix form as

$$p = Fa, \tag{4}$$

where p is an $IKGL$-element column vector of modeled dynamic projection data values, F is a matrix operator of dimension $IKGL \times (MNQR)$ containing the elements

$$f_{ikgl(mnq)} = \int_{t_k - \Delta t_k}^{t_k} U^m(d_i(t),t) V^n(t) W^q(\tau_g(t)) R^r(\xi_l(t)) dt \tag{5}$$

Here, a is an $MNQR$-element column vector of coefficients. The solution for the coefficients a in (4) can be estimated using maximum likelihood or Bayesian formulation.

The matrix operator F is a linear transformation from the image space with elements a into the projection space with elements p, which, in principle, includes effects of all physical processes such as attenuation, scattering and geometric point response of the collimator.

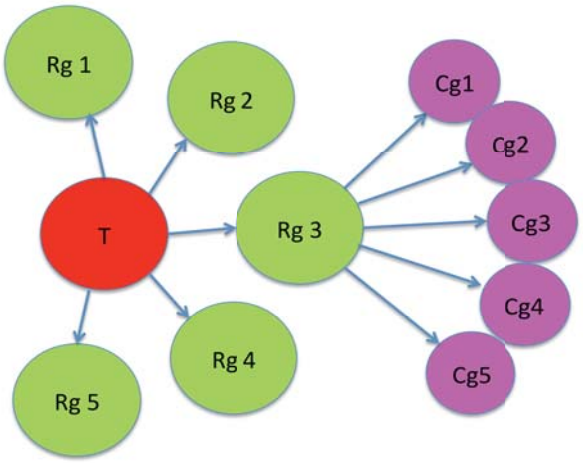


Fig. 2. Tree diagram representing temporal (T), respiratory (Rg), and cardiac gate (Cg) grouping for implementing the tensor product state (equation (1)) in the simulation.

III. SIMULATION METHODS

The mathematical cardiac torso (MCAT) phantom [14] was used to simulate the time-varying distribution of the radiopharmaceutical and motion of the heart due to respiration and cardiac beating. Time activity curves (TACs) for vital organs such as heart, lungs, liver, spleen, and several others were measured to simulate activity corresponding to a patient infused with the radiotracer ^{99m}Tc-tetrofosmin (^{99m}Tc, 140 keV) in our ongoing dynamic cardiac SPECT studies. Using these activities as input parameters, the noise-free projection data sets were generated for the SPECT geometry of a commonly available single-head gamma camera. The beating of the heart along with the respiratory and diaphragm motion was incorporated in the simulation. The period of cardiac cycle was 1 sec, and each systolic phase was accompanied after 0.325 sec of the end-diastolic phase. The respiratory period was 5 sec and the maximum displacement of the heart due to diaphragm motion was 2 cm. The camera rotation period was 15 sec such that there were 3 complete respiratory and 15 heart cycles for each rotation. In the simulation, we considered 5 respiratory and 8 cardiac gates, and 120 frames per rotation.

Although we simulated the activity over the whole torso, we particularly paid attention to the temporal evolution of the tracer diffusion in the myocardium. We considered the three-dimensional (3D) image space voxelized into 128×128×128 voxels of size 4.4 mm. A continuous spatial distribution using B-spline basis functions will delineate a smooth variation of the activity in space, and will be considered in future publications. Each detector was pixelized to a dimension of 128×128 with a pixel size of 4.4 mm. Each projection image was attenuation corrected while scatter and geometric point responses were ignored in the simulation.

We modeled the temporal dynamics of the tracer by modified B-spline basis functions that optimally capture the TACs for vital organs such as heart, lungs, and liver (see Ref. [19] for details). Although the beating of the heart is a continuous process, we assumed that the heart goes through a series of discrete cardiac states, as does the respiratory cycle. In order to obtain optimal temporal basis functions, we first fitted the TACs of the measured data. The basis functions for the cardiac and respiratory phases were chosen as Gaussians whose peaks corresponded to the discrete states in the cardiac coordinate. The widths of the functions were free parameters, and were tuned to maximally reflect the true motion of the heart.

In Fig. 2, we show the tree diagram for the 6D image reconstruction. As discussed in the Section II, we assumed the order of the gating of the projection data followed a cardiac (Cg) - respiratory (Rg) -temporal (T) sequence.

IV. RESULTS AND DISCUSSIONS

Figure 3 shows sampled reconstructed images to demonstrate the temporal evolution of tracer for the initial 3 minutes of infusion. There are particularly three different time scales: time for tracer uptake and wash out, the period of the respiratory cycle, and the period of the beating of the heart. In (A), we show the snapshots of the heart states for every 30 second time span to demonstrate the tracer kinetics. In (B) and (C), we show the heart motion due to respiration and cardiac beating for 3 sec and 1 sec time period respectively. The image intensities are in arbitrary scale. Since the respiratory and cardiac motions are coupled, both the displacement and the beating of the heart are visible in (B) and (C). In particular, the reconstruction clearly demonstrated the thickening of the myocardium during systole and widening of the left ventricle during diastole. The displacement of the myocardium in the vertical direction may be accounted for the motion due to respiration.

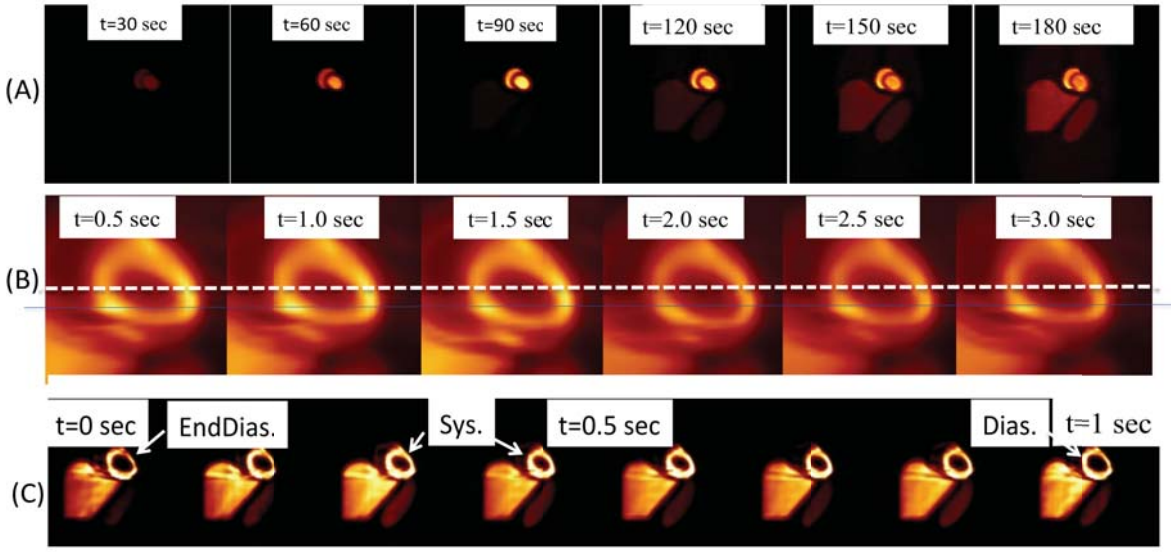


Fig. 3. Reconstructed images of the cardiac phases of the heart using fully 6D model demonstrating (A) temporal, (B) respiratory, and (C) cardiac motion. (A) Snapshots of the image at every 30 sec of the tracer infusion. (B) Vertical displacement of the endocardium due to respiration shown by thin line. (C) Heart state in one complete beat cycle, end-diastole (EndDias.) –Systole (Sys.)-Diastole (Dias.).

V. Conclusions

We have simulated the temporal evolution of the radionuclides using dynamic SPECT in connection with myocardial perfusion imaging when the motion due to beating of the heart and respiration are explicitly incorporated. A six-dimensional (6D) multiresolution spatiotemporal parameterization of the gated cardiac and respiratory data delineated the changes in the deformation of the heart as well as changes in the intensity of the signal caused by uptake and washout of the radiopharmaceutical in the myocardium. We believe our approach permits direct 6D quantitative visualization of the tracer distributions in moving organs and thus an accurate estimation of the parameter values that one can derive from the projection data set. Since the simulated parameters have been derived from a commonly available SPECT camera in the market, a future possibility of implementing the model in real patient data cannot be ruled out.

Acknowledgment

The authors would like to thank Paul Segars, Ph.D. and Benjamin M. W. Tsui, Ph.D. for use of the MCAT phantom.

References

[1] G. T. Gullberg, B. W. Reutter, A. Sitek, J. S. Maltz, and T. F. Budinger, "Dynamic single photon emission computed tomography-basic principles and cardiac applications," *Phys. Med. Biol.*, vol. 55, pp. R111-R191, Oct 21 2010.

[2] N. Matsumoto, D. S. Berman, P. B. Kavanagh, J. Gerlach, S. W. Hayes, H. C. Lewin, J. D. Friedman, and G. Germano, "Quantitative assessment of motion artifacts and validation of a new motion-correction program for myocardial perfusion SPECT," *Journal of Nuclear Medicine*, vol. 42, pp. 687-694, May 2001.

[3] A. Sitek, G. T. Gullberg, and R. H. Huesman, "Correction for ambiguous solutions in factor analysis using a penalized least squares objective," *IEEE Trans. Med. Imag.*, vol. 21, pp. 216-225, 2002.

[4] V. J. Cunningham and T. Jones, "Spectral-analysis of dynamic PET studies," *Journal of Cerebral Blood Flow and Metabolism*, vol. 13, pp. 15-23, Jun 1993.

[5] J. S. Maltz, "Parsimonious basis selection in exponential spectral analysis," *Phys. Med. Biol.*, vol. 47, pp. 2341-2365, Jul 7 2002.

[6] M. Unser, A. Aldroubi, and M. Eden, "Fast B-spline transforms for continuous image representation and interpolation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 277-285, Mar 1991.

[7] A. Sitek, R. H. Huesman, and G. T. Gullberg, "Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud," *IEEE Trans. Med. Imag.*, vol. 25, pp. 1172-1179, Sep 2006.

[8] R. M. Lewitt, "Multidimensional digital image representations using generalized Kaiser-Bessel window functions," *Journal of the Optical Society of America a-Optics Image Science and Vision*, vol. 7, pp. 1834-1846, Oct 1990.

[9] L. A. Shepp, Y. Vardi, J. B. Ra, S. K. Hilal, and Z. H. Cho, "Maximum-likelihood PET with real data," *IEEE Tran.Nucl. Sci.*, vol. 31, pp. 910-913, 1984.

[10] B. W. Reutter, G. T. Gullberg, and R. H. Huesman, "Direct least-squares estimation of spatiotemporal distributions from dynamic SPECT projections using a spatial segmentation and temporal B-splines," *IEEE Trans. Med. Imag.*, vol. 19, pp. 434-450, May 2000.

[11] J. Verhaeghe, Y. D'Asseler, S. Staelens, S. Vandenberghe, and I. Lemahieu, "Reconstruction for gated dynamic cardiac PET imaging using a tensor product spline basis," *IEEE Trans. Nucl. Sci.*, vol. 54, pp. 80-91, Feb 2007.

[12] B. Feng, P. H. Pretorius, T. H. Farncombe, S. T. Dahlberg, M. V. Narayanan, M. N. Wernick, A. M. Celler, J. A. Leppo, and M. A. King, "Simultaneous assessment of cardiac perfusion and function using 5-dimensional imaging with Tc-99m teboroxime," *J. Nucl. Cardiol.*, vol. 13, pp. 354-361, May-Jun 2006.

[13] U. Shrestha, F. Alhassen, et al, "Reconstruction of gated dynamic cardiac SPECT data using spatiotemporal basis functions," in 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record, pp. 2449–2492, Sept 2012.

[14] W. P. Segars and B. M. W. Tsui, "MCAT to XCAT: The Evolution of 4-D Computerized Phantoms for Imaging Research," *Proceedings of the IEEE*, vol. 97, pp. 1954-1968, Dec 2009.

Data Correction Methods for Wholebody Patlak Imaging from List-mode PET Data

Wentao Zhu, Bing Bai, Peter S. Conti, Quanzheng Li, and Richard. M. Leahy

***Abstract*—We previously described a new approach using dual-time-point list-mode PET data to perform Patlak analysis. This approach can be used to compute wholebody Patlak parametric images from a two-pass PET scan. Our method directly fits the Patlak slope and intercept at each voxel using the list mode arrival time for each event and a penalized maximum likelihood estimator. In order to make this method feasible for routine clinical studies, we need to address the following important issues in patient data processing: dynamic modeling of random and scattered events, estimation of the blood input function, and compensation for patient movement between frames. We describe how we handle each of these issues. We then illustrate the approach in application to wholebody FDG studies and compare our results to those obtained with standard static imaging and with fractional SUV methods.**

***Index Terms*—PET, Patlak, listmode, wholebody**

I. Introduction

STATIC positron emission tomography (PET) scan and semi-quantitative measurements such as the standardized uptake value (SUV) are widely used in cancer staging and in following response to therapy. SUV does not reflect the underlying dynamics of tracer uptake since it depends on uptake integrated over a single frame. Variations in protocols among clinical sites and the complex relationship between dose, uptake and body weight also limit its use as a quantitative biomarker.

To overcome these limitations others have investigated the use of dual-time-point PET scans. Studies have shown that this approach can improve diagnostic accuracy for certain cancers such as malignant lung nodules [1] and high-grade brain tumors [2] when comparing to the standard single frame SUV method. It is also useful in differentiating malignant lesions from benign ones [3]. However, dual-time-point SUV is still affected by some of the factors mentioned above and remains at best a semi-quantitative measure.

This work is supported by NIH grants R01EB013293 and R01EB010197.

Wentao Zhu is with the University of Southern California, LA, CA 90089 USA (e-mail: wentaozh@usc.edu).

Bing Bai is with the University of Southern California, LA, CA 90089 USA (e-mail: bbai@usc.edu).

Peter S. Conti is with the University of Southern California, LA, CA 90089 USA (e-mail: pconti@hsc.usc.edu).

Quanzheng Li is with Massachusetts General Hospital, Boston, MA, 02114 USA (e-mail: li.quanzheng@mgh.harvard.edu)

Richard M. Leahy is with the University of Southern California, LA, CA 90089 USA (e-mail: leahy@usc.edu).

In contrast, pharmacokinetic models based on dynamic PET data offer the ability to fit quantitatively meaningful parameters that reflect underlying metabolic or other physiological processes. This in turn can lead to more accurate detection and staging of tumors as well as differentiation from other forms of enhanced uptake, such as regions of inflammation. Freedman et al. [4] found that discrepancies between SUV and Patlak values in patient therapy are primarily due to the unmetabolized FDG measured by SUV, which is important for the accurate determination of glucose metabolic rates. They concluded that the Patlak slope is more accurate in predicting response to treatment. Graham et al. [5] also concluded that Patlak slope is a better discriminator between normal tissue and tumors.

Clinical PET systems have limited axial extent that precludes wholebody dynamic scanning. Typically the patient is scanned for a few minutes at several contiguous bed positions and the reconstructed images are stitched together to form a wholebody image. If we want to perform dynamic analysis while still collecting wholebody data we can only sample part of the dynamic process and must then fit the dynamic model with only partial data. The Patlak model has only two parameters and does not directly use the earlier part of the time activity curve. We can therefore compute these parameters from measurements collected over two time intervals or frames, provided sufficient time elapses between these frames.

Motivated by the goal of performing wholebody Patlak imaging we have developed a Patlak estimation method from dual-time-point list-mode data [6]. Simulation and Cramer-Rao analysis in [6] suggest that this method can achieve superior differentiation of tumor from background in small tumors compared to using fractional changes in SUV computed from the same dual-time-point data. To apply this approach to clinical data we need to account for the time varying effect of random and scattered events and also for patient motion. In Section 2 we describe how we model these factors. We also describe how we estimate the input function necessary to perform Paltak analysis. Finally we apply our method to wholebody patient 18F-FDG data, and compare the Patlak image derived using our method with single and dual frame static PET images.

II. Method

2.1 Patlak Estimation from two Frame List Mode Data:

The Patlak graphical model applies to kinetic data beyond time $t \geq T^*$ at which point changes are effectively due to irreversible trapping in a single compartment. Let $\eta(t)$ be the

tracer time activity curve (TAC) with input function $C(t)$. We can write the Patlak equation as:

$$\eta(t) = \kappa \int_0^t C(\tau)d\tau + qC(t)$$

where κ is the net influx rate, and q is the intercept of the Patlak model. We can therefore model the rate function at voxel *j* after steady state $t \geq T^*$ as a linear combination of two basis functions $B_1(t)$ and $B_2(t)$:

$$\eta_j(t) = \sum_{l=1}^{2} w_{jl} B_l(t)$$

$$\omega_{j1} = \kappa_j, \omega_{j2} = q_j, B_1(t) = \int_0^t C(\tau)d\tau\,, B_2(t) = C(t)$$

Consequently the rate function in sinogram space at line of response (LOR) *i* can be written as:

$$\lambda_i(t) = e^{-\frac{t}{\tau}} \sum_{j=1}^{n_v} \sum_{l=1}^{2} p_{ij} w_{jl} B_l(t) + r_i(t) + s_i(t)$$

where p_{ij} is the probability of an event at voxel *j* being detected at detector pair *i*, and n_v the number of voxels; the exponential term accounts for radioactive decay. Time series $r_i(t)$ and $s_i(t)$ are the random and single rates for LOR *i* at time *t*. Assuming we have list mode data over the interval $[t_1\ t_2]$ and $[t_3\ t_4]$, and that the arrival times follow the time-inhomogeneous Poisson process $\lambda_i(t)$, we estimate the Patlak parameters in a Bayesian framework. The log posterior function is:

$$F(W) = -\sum_{i=1}^{n_p} \sum_{k=1}^{x_i} log\lambda_i(a_{ik}) + \sum_{i=1}^{n_p} \left(\int_{t_1}^{t_2} \lambda_i(t)dt + \int_{t_3}^{t_4} \lambda_i(t)dt\right) + \beta \frac{1}{2} w^T R_S w$$

where a_{ik} is the arrival time of the k^{th} event at LOR *i*, and R_S is the 2nd derivative matrix of a quadratic spatial penalty (or equivalently the 2nd derivative of a quadratic Gibbs energy function). We do not regularize in time since we only have 2 basis functions. We maximize $F(W)$ to compute estimates of the Patlak parameters at each voxel.

2.2 Randoms and scatters in dynamic list-mode data

Since we are using a list mode model, we need a continuous-time estimate of both randoms and scatters. As our protocol is composed of 2 listmode acquisition frames each with 5 bed positions, in preprocessing we obtained 10 pairs of static randoms and scatters estimates from 10 listmode files. We make the simplifying assumption that the temporal and spatial distribution of scatters and randoms are separable [7]. Let R_{bi} denote the total number of delayed events for the i^{th} LOR integrated over the entire acquisition period of the b^{th} bed position. Let $g_b(t)$ be an estimate of the dynamic variation of randoms averaged over all LORs and normalized to integrate to unity over the acquisition time of the b^{th} bed position. Then the estimated randoms rate function in the b^{th} bed position can be modeled as:

$$r_{bi}(t) = R_{bi} g_b(t) \quad (8)$$

Similarly, let S_{bi} denote the total number of scattered events at the *i*th LOR in the b^{th} bed position and $h_b(t)$ be a normalized estimate of the dynamic variation, then the scatter rate function in the b^{th} bed position can be modeled as:

$$s_{bi}(t) = S_{bi} h_b(t) \quad (9)$$

In the following, the integrated randoms R_{bi} are computed from a sinogram of delayed events summed over the acquisition period of the b^{th} bed position. Integrated scatters S_{bi} is computed from the sinograms generated by histogramming the list mode data of the b^{th} bed position, in combination with a co-registered x-ray CT scan, using a deterministic scatter estimation method [8]. The temporal rate function of scatters for each bed position is assumed to be proportional to the rate function of prompts, and then normalized to unity over the frame, i.e.

$$h_b(t) = \frac{p_b(t)}{\int_{L_b}^{U_b} p_b(t)dt} \quad (10)$$

where L_b and U_b are the starting and end time of the b^{th} bed. The rate function of all prompts for each bed is obtained by first dividing the 5 min time into ten 30 sec subsets and counting the number of prompts for each subset, and then performing interpolation between the 10 data points. A similar process is used for estimating the randoms rate function $g_b(t)$, but since these vary as the square of activity we use:

$$g_b(t) = \frac{{p_b}^2(t)}{\int_{L_b}^{U_b} {p_b}^2(t)dt} \quad (11)$$

2.3 Blood input function

Our goal in this work is to develop a practical method for wholebody clinical Patlak studies. Since routine measurement of the input function using arterial blood sampling is impractical, here we adopt a hybrid approach that combines a population based approach for $t < t_1$ with a simplified exponential model for $t > t_1$.

We pooled data from a set of N=7 60min FDG dynamic scans of patients with liver cancer. We identified ROIs in the abdominal aorta in each subject from coregistered contrast enhanced CT images. We extracted estimates of activity in the aorta from MAP reconstructions [9] of each frame to fit a continuous model. We then scaled each input function to have unit integral.

$$C^i_{scale}(t) = \alpha_i C^i_{origin}(t) \qquad i \in 1,2,\dots N$$

$$\alpha_i \int_0^T C^i_{origin}(t)dt = 1$$

The population average was then generated by averaging over all scaled input functions, i.e.

$$C_{standard}(t) = \frac{1}{N} \sum_{i=1}^{N} C^i_{scale}(t)$$

We then adopt the following input function model:

$$C(t) \triangleq \begin{cases} \mu C_{standard}(t) & t < t_1 \\ \beta\, exp(\gamma t) & t \geq t_1 \end{cases}$$

where t_1 is the scan start time, which is approximately 1 hour after injection.

The constants μ, β, γ are chosen so that the input function is continuous at t_1. To find these constants we choose the two frames corresponding to the bed position that contains the abdominal aorta. For each frame we compute a static MAP reconstruction. An ROI within the aorta is manually delineated and the average activity in the ROI computed for both frames.

The parameters μ, β, γ are then computed by fitting the model to these two values by integrating of their respective durations. Note that although there are three parameters, μ and $\beta$are dependent because of the continuity constraint. This approach could be modified in a straightforward manner to include data from additional frames that also include large arterial vessels.

2.4 Motion correction

Our motion correction technique performs pairwise nonrigid registration between each set of two frames that correspond to the same bed position. We first compute static MAP reconstructions for each frame at each bed position using a spatially variant smoothing function selected to achieve a constant count independent resolution [10]. We then use a non-rigid mutual information based registration tool [11] to find the deformation field that coregisters each pair of images. Patlak images are then computed using this deformation field. Since we cannot warp the data for the 2nd frame, the reconstruction method is modified so that the spatial mapping between the 1st and 2nd frame is propagated through the forward and backprojection operators when computing the gradient of the cost function with respect to the 2nd frame.

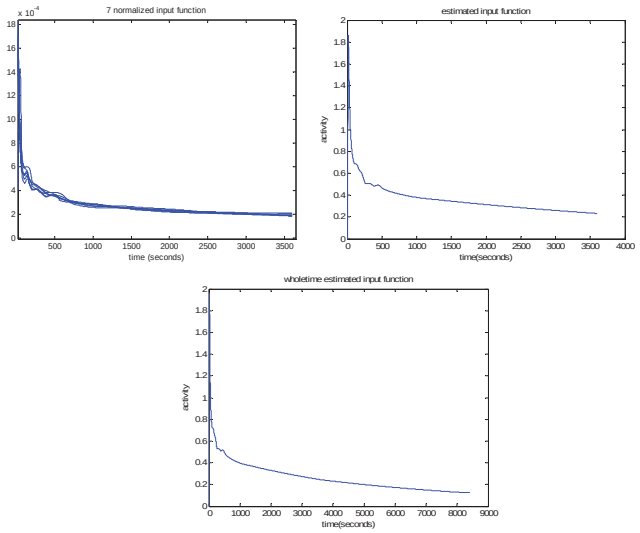


Fig. 1 (left) the input function from 7 liver scan patients. **(right)** the average input function from time 0 to 60min, and (**bottom**) estimated the input function formed by extrapolation of the population average beyond 60mins using data from the abdominal aorta.

III. Results and Discussion

3.1 Blood input function estimation

We used n=7 FDG liver scan data sets to train the input function. 60mins of data from the time of injection was collected and sorted into a total of 20 frames (6x30sec, 7x1min, 4x5min, 3x10min). Each frame was reconstructed with matched resolution and the activity in a hand-drawn ROI computed for each frame. A continuous time function was fitted using spline interpolation and the curve normalized to integrate to unity from 0 to 60 min. The 7 normalized input functions are shown in Fig. 1 (left). Shown in Fig. 1 (right) is the average input function from 0 to 60 min. We then extrapolated these curves using the exponential model described above for each wholebody subject separately. Fig. 1 (bottom) is the estimated input function using the exponential function approximation after 60 min for one of the whole body examples shown below.

3.2 18F-FDG wholebody patient study

We applied the dual-time-point list mode estimation method to wholebody patient data collected using a Siemens Biograph PET/CT scanner at the University of Southern California PET center. The 3D PET system's diameter is 855.20mm. The detector size is 2.673mm × 2.673mm × 2.025mm. The number of rings is 55 and the maximum ring difference is 38. A span of 11 resulted in a total of 559 sinograms with 336 angles of view by 336 radial LORs for each sinogram. List mode data was collected at 5mins per bed position starting at 45mins and 105mins post injection respectively. Static reconstructions at matched resolutions were reconstructed from each frame and the Patlak image reconstructed from both frames using the motion correction and other calibration methods described above. The images were then stitched to form the wholebody images shown below.

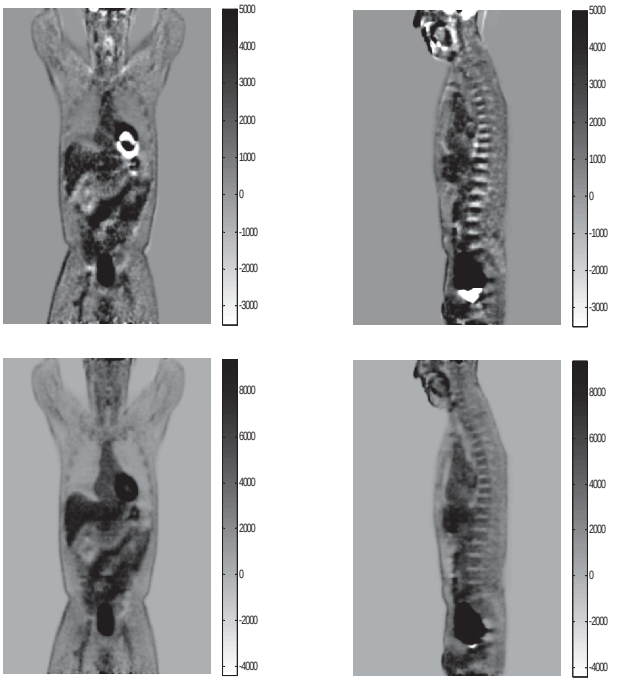


Fig. 2 Difference SUV images of frame 1 and frame 2 before (**top 2 images**) and after registration (**bottom 2 images**)

Fig 2 shows the difference SUV images (SUV2-SUV1) for one patient before and after registration. In this example the patient left the table between the two scans presenting a more difficult registration challenge than in the typical case. From the top two images in Fig. 2 we can clearly see the bright and dark regions caused by misregistration in the heart, brain, liver and vertebrae. The post-registration result at the bottom of Fig. 2 clearly shows significantly better alignment between the 2 frames. The same deformation field was used to compute the Patlak images.

Fig. 3 shows Patlak slope images with and without random and scatter correction and motion compensation. Overall, these corrections produce noticeable improvements in contrast and noise suppression as well as improved boundary definition as a result of coregistration. This patient had a metal implant in the left leg which produces a clear artifact since the corrections we used do not account for its highly scattering and attenuating properties.

For comparison we also computed a fractional SUV image as %DSUV = (SUV2-SUV1)/SUV1. Because the SUV normalization by weight and dose cancels in this ratio, we computed %DSUV directly from the MAP images. Since we computed these on a voxel wise basis, rather than over an ROI as would typically be done, the estimates are very noisy (column 4 in Fig. 4). While this would not be done in practice, the fact that the Patlak image quality is good indicates the potential of our approach to produce meaningful dynamic wholebody parameteric images in a manner that cannot be achieved using fractional SUVs.

Comparing the Patlak image and static PET images (either frame) reveals only subtle differences visually, as shown in Fig. 4. However it is important to note that the Patlak result is a voxel-wise quantitative image so that the numerical values themselves should be more informative for staging and treatment planning.

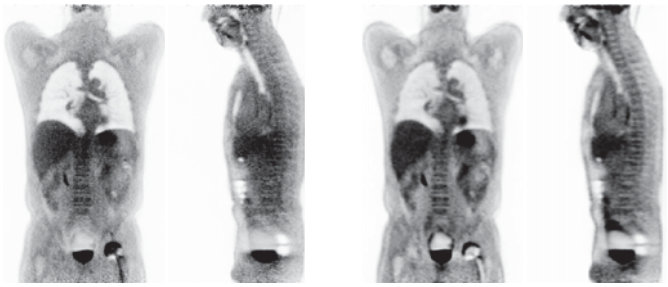


Fig. 3 Patlak slope images: (**left**) no scatter or random correction, no motion correction. (**right**) scatter, random correction, motion corrected.

IV. CONCLUSION

We addressed several practical issues in wholebody Patlak image estimation. We introduced a hybrid population-based method for estimating the blood input function. We also demonstrated the effectiveness of our random and scatter correction and of motion compensation. The wholebody patient data shows that we can produce Patlak images of equivalent quality to static PET scans. This is in contrast to dual-time point fractional SUV values which do not produce good quality images. Clearly further studies are required to determine whether this approach leads to improve diagnostic power relative to the alternatives. However, with the methods described about we are now in a position to perform these studies.

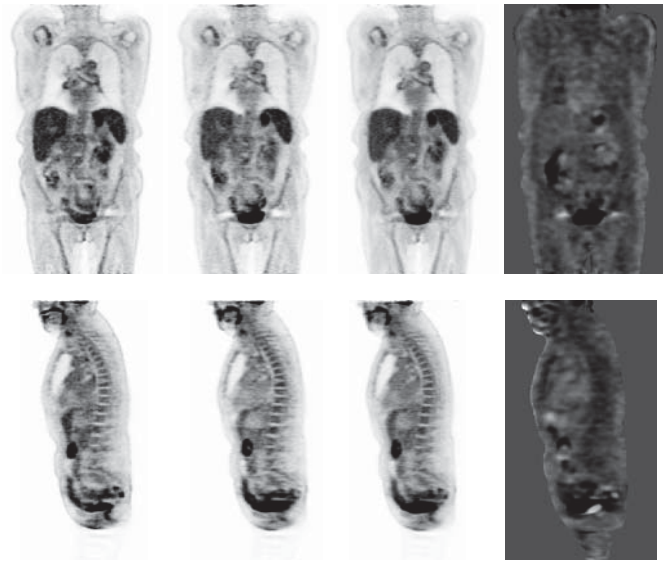


Fig. 4. First frame (**column 1**), second frame (**column 2**), Patlak (**column 3**) and %DSUV (**column 4**) wholebody reconstruction images.

REFERENCES

[1] K. Alkhawaldeh, et al., "Impact of dual-time-point 18F-FDG PET imaging and partial volume correction in the assessment of solitary pulmonary nodules, " *Eur J Nucl Med Mol Imaging*, vol. 35(2), pp: 246–252, 2008

[2] E. Prieto, et al., "Voxel-based analysis of dual-time-point 18F-FDG PET images for brain tumor identification and delineation, " *J Nucl Med*, vol. 52(6), pp: 865-872, 2011

[3] R. Hustinx, et al., "Dual time point fluorine-18 fluorodeoxyglucose positron emission tomography: a potential method to differentiate malignancy from inflammation and normal tissue in the head and neck, " *Eur J Nucl Med*, vol. 26(10), pp: 1345-1348, 1999

[4] N. M. T. Freedman, et al., "Comparison of SUV and Patlak slope for monitoring of cancer therapy using serial PET scans, " *Eur J Nucl Med Mol Imaging*, vol. 30(1), pp: 46-53, 2003

[5] M. M. Graham, et al., "Comparison of simplified quantitative analyses of FDG uptake, " *Nucl Med Biol*, vol. 27(7), pp: 647-655, 2000

[6] W. Zhu, Q. Li, R. M. Leahy. "Dual-Time-Point Patlak Estimation from List Mode PET Data," *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp: 486:489

[7] TE Nichols, J. Qi, E. Asma, RM Leahy, "Spatiotemporal Reconstruction of List Mode PET Data," *IEEE Trans Med Imag*, vol 21 (4), pp: 396-404 Apr 2002

[8] C.C. Watson, "New, faster, image-based scatter correction for 3D PET," *Nuclear Science, IEEE Transactions on*, vol. 47(4), pp: 1587-1594, 2000

[9] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou and T. H. Farquhar, "High Resolution 3D Bayesian Image Reconstruction Using the MicroPET Small-animal Scanner," *Physics in Medicine and Biology*, vol 43(4), pp: 1001-1013, 1998

[10] Q. Li, B. Bai, S. Cho, A. Smith, and R. Leahy, "Count independent resolution and its calibration," in Proceedings of 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, pp: 223–226, 2009.

[11] Dirk-Jan Kroon, registration toolbox "non-rigid b-spline grid image registration", MAY 2008.

Multi-dimensional sinogram restoration for myocardial blood flow estimation from dose-reduced dynamic CT

Dimple Modgil, Adam M. Alessio, Michael D. Bindschadler, Kevin J. Little, David Rigie, Philip A. Vargas and Patrick J. La Rivière

***Abstract*—Quantification of myocardial blood flow provides valuable diagnostic and prognostic information for the evaluation of coronary heart disease. Dynamic CT offers the potential of a quick and cost-effective method to quantify the myocardial blood flow. However, a primary reason dynamic CT is not widely employed is the large radiation dose imparted to the patient. Radiation dose reduction is essential for clinical acceptance of these studies. In order to achieve dose reduction, one needs to be able to extract the same blood flow information from much noisier data. We propose that it is possible to do so by applying sinogram restoration techniques to dynamic CT data. So far, these techniques have only been applied to static data. We expand this technique to also impose regularization in the temporal domain. In this paper, we present preliminary results using a variety of temporal regularization techniques on noisy, simulated dynamic CT data.**

***Index Terms*—Dynamic CT, sinogram restoration, coronary heart disease, myocardial blood flow, time attenuation curve**

I. Introduction

Cardiac diagnostics would benefit greatly from an easy and widely available technique for quantitative myocardial blood flow estimation. Quantification of myocardial blood flow (MBF) is clinically very important for evaluation of coronary heart disease (CHD). The most common diagnostic test for CHD using angiography is insufficient for clinical decision making. Studies have shown that perfusion information leads to better outcomes and reduced costs [1]. Unlike modalities such as single photon emission computed tomography (SPECT), which can only detect unbalanced ischemia, quantification of MBF can assess triple vessel disease. While the absence of ischemia is a good predictor of event-free survival, the severity of ischemia, as measured with quantitative MBF, is closely correlated to the occurrence of adverse events further motivating quantification of MBF. Dynamic CT offers a quick method to measure MBF in absolute units (ml/g/min), but it has a big drawback, namely, the large radiation dose that is received by the patient during this test. CT manufacturers have primarily focused their research and development on the mainstream applications of static imaging with high anatomic detail. As such, CT hardware and software are optimized for high-flux, high throughput imaging, with little optimization for 4D imaging of the time course of contrast agents.

Adam Alessio and Michael Bindschadler are in the Department of Radiology at the University of Washington. The other authors are with the Department of Radiology, The University of Chicago.

There are two potential methods to reduce dose from dynamic CT scans: reduced tube current or reduced time samples or a combination of both. Both these strategies come at the cost of increased noise while reducing dose. Several groups have explored techniques to incorporate temporal variation in cardiac imaging, for a low dose scan protocol. For example, Sawall *et al.* [2], [3] used a variant of McKinnon-Bates image reconstruction algorithm along with bilateral filtering in multiple dimensions (three spatial, three temporal: cardiac, respiratory and perfusion) to achieve low-dose phase-correlated imaging in small animals. Ritschl *et al.* [4] used a new method for spatial and temporal regularization for temporal-correlated CT image reconstruction. Their method utilized total variation constraint in both spatial and temporal dimensions. Chen *et al.* [5] used a prior image constrained compressed sensing algorithm in the context of time-resolved cardiac C-arm cone-beam CT.

In this work, we explore 4D sinogram restoration methods to restore noisy CT data so as to minimize the radiation dose from dynamic CT. Ultimately, we want to verify that accurate and precise time attenuation curves and parametric estimates can be recovered from CT acquisitions with substantially lower dose to patients.

II. Methods

Noise in CT data is generally controlled through the apodization of the kernel used in the filtered backprojection reconstruction. This simple approach does not model the spatially variant noise properties inherent to the data and is suboptimal. Iterative sinogram restoration has been proposed to improve the signal-to-noise ratio of CT data [6], [7]. These restoration strategies generally model several of the degrading physical effects of the CT acquisition and attempt to remove these degradations from the data through an iterative deconvolution based on a reasonable stochastic model for the data and a simple roughness-penalty term. These methods offer the benefits of both noise and artifact reduction. In this paper, we propose tailoring these approaches for use with dynamic CT data. We use the same general model that we have used for static sinogram restoration [6], with a slightly different interpretation of the terms. We assume the raw data are modeled as:

$$Y_i = \int \text{Poisson}\left\{I_0(E)\exp\left(-l_i\right)\right\} EdE + \text{Normal}\{0,\sigma_i^2\} \quad (1)$$

$$l_i \equiv \int_{L_i} \mu(\mathbf{x},E)dl. \quad (2)$$

This represents the sum of a compound Poisson distribution and a Gaussian electronic noise term. Here, Y_i are the measurements recorded by an energy-integrating detector at various bins, views and time frames, L_i is a line through the object for the ith measurement and $I_0(E)$ is the incident polyenergetic X-ray spectrum. The total number of measurements now is the product of the number of projection angles, the number of detector bins and the number of temporal frames. This model takes into account the polychromatic X-ray spectrum, compound Poisson noise and electronic Gaussian noise. To yield a tractable likelihood function, we model the first term as a scaled, over-dispersed Poisson likelihood and add to it a constant derived from the electronic noise and dark current. The distribution of the resulting shifted Poisson variable matches the true distribution in the first and second moments [6]. The goal is to estimate the set of line integrals, l_i (which are a function of projection bin, azimuthal angle, and temporal frame), as accurately as possible from noisy data. We do so through a penalized likelihood strategy tailored for the dynamic CT application. We add to the cost function a penalty term that will penalize the temporal variations in the sinogram due to iodine transport. The general form of the objective function is:

$$\hat{l} = argmax_{p\geq 0}\left[L(l;y) - \beta R(l) - \gamma C(l)\right] \quad (3)$$

where the first term is the usual Poisson log-likelihood function, the second is a spatial roughness penalty, and the third term is a temporal roughness penalty. We will try several different penalty terms in the temporal domain:

1) $\gamma = 0$: smoothing only in the spatial domain.
2) $\beta = \gamma$: common strength for quadratic penalty over both spatial and temporal domains.
3) $\beta \neq \gamma$: different strength for quadratic penalty over spatial and temporal domains.
4) $\gamma = \hat{\gamma}(t)$: strength of temporal penalty is a function of time to allow for variation in bias to variance tradeoff at different times in the acquisition. This may be beneficial considering rising edge of iodine enhancement informs the blood flow estimate more than tail and has more signal change than the tail. This work will be investigated later.
5) Temporal shape-based penalty: This will encourage adherence of the temporal variation of the iodine enhancement to the priors term that contains a priori knowledge of the temporal variations due to iodine transport. This constraint will encourage, but not require, adherence of the reconstructed time attenuation curves to the priors. This work will be investigated later.

A. Smoothing only in the spatial domain

In the first case, we have no temporal smoothing term and a quadratic smoothing penalty term in the spatial domain. The objective function is given by [6]:

$$\hat{l} = argmax_{l\geq 0}\left[L(l;y) - \beta R(l)\right], \quad (4)$$

$$R(l) = \sum_{k=1}^{K}\psi_k\left(\sum_{j=1}^{N_p} t_{kj}\right) \quad (5)$$

$$\psi_k(x) \equiv \frac{x^2}{2} \quad (6)$$

The inner sum over j involves the difference between the kth projection bin and its nearest neighbors. In this work, we chose t_{kj} it to create pairwise combinations among a given projection bin and its four nearest neighbors in the spatial domain. Following the separable parabolic surrogate method (SPS) to maximize the likelihood, as derived by Ergodan et al. [8], and the approach by La Rivière et al. [6], the update equation is given by:

$$l_i^{(n+1)} = l_i^{(n)} + \frac{\dot{h}_i(l_i^{(n)}) - \beta\sum_{k=1}^{K}\sum_{j=1}^{N} t_{ki}t_{kj}l_j^{(n)}}{c_i^{(n)} + \beta\nu_i}, \quad (7)$$

$$l_i \equiv f(l_i^{mono}) \quad (8)$$

$$\dot{h}_i(l) = I_i e^{-l}\left[1 - \frac{y_i}{I_i e^{-l} + r_i}\right], \quad (9)$$

$v_i \equiv \sum_{k=1}^{K}|t_{ki}|t_k$, with $t_k \equiv \sum_{i=1}^{N_Y} t_{ki}$, and the $c_i^{(n)}$ are the curvatures of the paraboloidal surrogates [6].

B. Common strength for quadratic penalty in the spatial and temporal domains

In this case the objective function is given as above except for a difference in the pairwise combinations:

$$\hat{l} = argmax_{l\geq 0}\left[L(l;y) - \beta R'(l)\right], \quad (10)$$

$$R'(l) = \sum_{k=1}^{K}\psi_k\left(\sum_{j=1}^{N_p} t'_{kj}\right) \quad (11)$$

$$\psi_k(x) \equiv \frac{x^2}{2} \quad (12)$$

The inner sum over j involves the difference between the kth projection bin and its nearest neighbors. In this work, we choose t'_{kj} it to create pairwise combinations among a given projection bin and its four nearest neighbors in the spatial domain and two nearest neighbors in the temporal domain.

C. Different strength for quadratic penalty in the spatial and temporal domains

In this case the objective function is given by:

$$\hat{l} = argmax_{l \geq 0} \left[L(l; y) - \beta R(l) \right] - \gamma C(l), \tag{13}$$

$$C(l) = \sum_{k=1}^{K} \psi_k \left(\sum_{j=1}^{N_p} t''_{kj} \right) \tag{14}$$

$$\psi_k(x) \equiv \frac{x^2}{2} \tag{15}$$

The spatial penalty term is the same as in section 2.1. In the temporal penalty term, the inner sum over j involves the difference between the kth projection bin and its nearest neighbors. In this work, we chose t''_{kj} to create pairwise combinations among a given projection bin and its two nearest neighbors in the temporal domain. For the spatial penalty term, we choose t_{kj} to create pairwise combinations among a given projection bin and its four nearest neighbors in the spatial domain.

D. Data Generation and Image Reconstruction

We generated dynamic material phantoms to mimic the exchange of iodine in the myocardium. Using the XCAT phantom as the base image volume, we simulated 4D image volumes with unique kinetics in the right ventricular cavity, left ventricular cavity, aorta, and myocardium. These kinetics were derived from our detailed mathematical model of iodine exchange. Starting from a ventricular or aortic input signal, the model accounts for the time delay and signal dispersion in large vessels between the input measurement and myocardium. In the tissue region of interest, the model accounts for flow heterogeneity and axial dispersion in the concentration of iodine as it exchanges with the interstitial space. In addition, the known differences between the systemic and microvascular hematocrits are appropriately handled. This iodine exchange model was informed by patient and porcine studies and allows for generation of realistic iodine curves for a range of physiologic states. These curves defined the changes in the dynamic phantom, which consisted of concentration maps of three materials (water, bone and iodine), as shown in figure 1, for each time frame.

Figure 1. A slice from one frame of the dynamic CT phantom consisting of iodine, bone and water

We constructed a polyenergetic simulator for CT projection data that uses the system model described above by eq. 1. The spectrum and the system parameters used by our simulator matched a CT scanner at the University of Washington (GE Lightspeed 16-slice, GE Healthcare, Waukesha, WI). Noisy projection data was generated with a tube current of 25 mA for this phantom for 25, 1 sec time frames. The noisy projection data was then processed through our proposed sinogram restoration methods. Beam hardening correction was computed for this spectrum assuming a uniform water phantom at a monochromatic average energy of 62.33 keV. This correction was applied to the restored sinogram data. The images were reconstructed from the restored sinograms using filtered backprojection (FBP) with a Hann filter with a cutoff of 0.46. We also reconstructed the images from the noisy, non-restored sinogram data with FBP.

In the images for each frame, a mean value of Iodine concentration was computed in four different ROIs in the left myocardium. From this data, a plot of variation in iodine concentration versus time was obtained. This curve is referred to as a time attenuation curve (TAC). The shape and magnitude of these curves can be used to estimate MBF using models of iodine transport in the myocardium [9], [10].

III. Results

Figure 2 shows the heart region for images reconstructed using both unrestored and restored sinogram data. Images reconstructed using unrestored sinogram data are much more noisy. The four ROIs chosen for computing the TACs were placed in the wall region of the left ventricle.

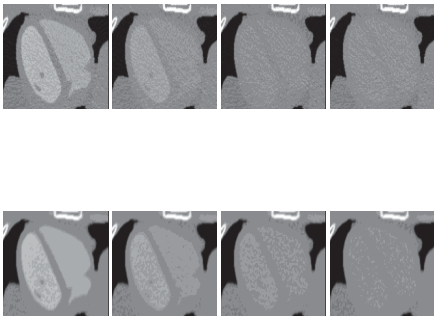


Figure 2. Reconstructed image slice, showing the heart region, for time frames at 6, 11, 16 and 21s: top- using unrestored sinogram data, bottom - using restored sinogram data with $\beta = 1000, \gamma = 500$

The time attenuation curves are presented in figure 3. The number of frames used for reconstruction were: 25, 12 and 8, to represent three potential acquisition schemes acquiring an image every second, every other second (1/2 dose), and every 3 seconds (1/3 dose). In this figure, the legend refers to the following methods used for image restoration:

1) No restoration: FBP reconstruction with no sinogram restoration
2) beta=1000, no temporal: sinogram restoration only in spatial domain ($\beta = 1000$) followed by FBP
3) beta=gamma=1000: sinogram restoration with same quadratic penalty in both spatial and temporal domains followed by FBP
4) beta=1000, gamma=500: sinogram restoration with different quadratic penalty terms both spatial and temporal domains followed by FBP

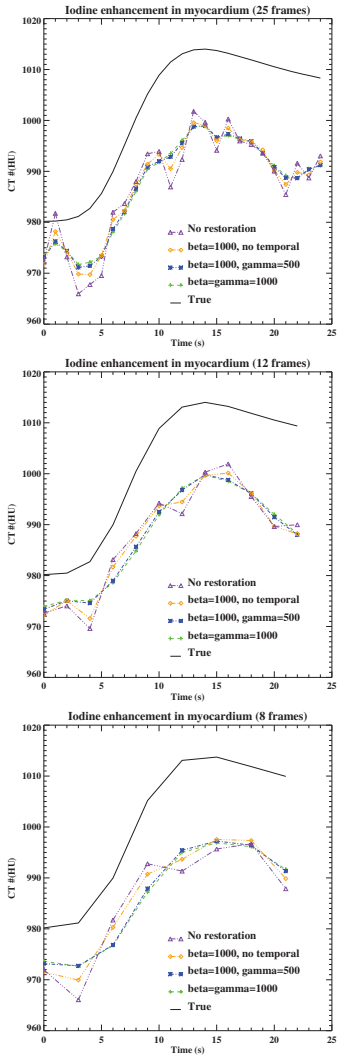


Figure 3. Time attenuation curves obtained using sinogram data from 25, 12 and 8 frames

The enhancement as measured in the reconstructed images is less than the true value due to beam hardening. Currently, our beam hardening correction only corrects for water. We will add multi-material beam hardening correction in the future. We observe from the figure that the simulated TAC curves mimic very closely the true TAC curve when all the time frames (25) are used. The simulated curves reconstructed using smoothed sinogram data match much better than the unsmoothed one. Looking at the undersampled data using 12 frames, we observe that the smoothed data with temporal smoothing follows the true curve much more closely. This difference is even more obvious when we use 8 frames for reconstruction. We do not see any perceptible difference between performing the temporal smoothing separately from spatial smoothing versus performing the smoothing in both domains together.

IV. Conclusions and Further Work

We found that sinogram restoration in the spatial and temporal domains reduces the variance in the time attenuation curves, especially in undersampled temporal data. We found that the slope of the curve, which is a predictor of flow, is closer to the true slope when sinogram restoration is performed with spatial and temporal regularization.

In the future, we will explore the use of shape-based or frequency-informed temporal priors in the temporal smoothing. There are a variety of ways of obtaining suitable temporal priors. We will also explore the use of the Karhunen-Loeve (KL) transform of a sample covariance matrix calculated for the portion of the sinogram data that is the forward projection of a region of interest tightly encircling the heart. Prior work with dynamic PET reconstruction has shown that there are usually only a few significant temporal components in the expansion [11]. These will be used as the spatially-variant priors in the above cost function. Unlike the work of Kao et al., which explicitly discards all other temporal components, the cost-function based approach will only encourage adherence to the priors while allowing the data to dictate other temporal variations.

V. Acknowledgements

This work was supported by NIH grants R01HL109327, R01CA134680 and R21EB009450. We thank Drs. Ting Lee and Aaron So for access to dynamic patient and porcine studies for developing the exchange model.

References

[1] P.A. Tonino, B. De Bruyne B, N.H. Pijls et al., “Fractional flow reserve versus angiography for guiding percutaneous coronary intervention”, N. Engl. J. Med., vol. 360, pp. 213-224, 2009.

[2] S. Sawall, F. Bergner, R. Lapp, M. Mronz, M. Karolczak, A. Hess and M. Kachelriess, “Low-dose cardio-respiratory phase-correlated cone-beam micro-CT of small animals, Med. Phys., vol. 38, no. 3, pp. 1416-1424, 2011.

[3] S. Sawall, J. Kuntz, M. Socher, M. Knaup, A. Hess, S. Bartling and M. Kachelriess, “Imaging of cardiac perfusion of free-breathing small animals using dynamic phase-correlated micro-CT”, Med. Phys., vol. 39, no. 12, pp. 7499-7506, 2012.

[4] L. Ritschl, S. Sawall, M. Knaup, A. Hess and M. Kachelriess, “Iterative 4D cardiac micro-CT reconstruction using an adaptive spatio-temporal sparsity prior”, Phys. Med. Biol., vol. 57, pp. 1517-1525, 2012.

[5] G-H. Chen, P. Teirault-Lauzier, J. Tang, B. Nett, S. Leng, J. Zambelli, Z. Qi, N. Bevines, A. Raval, S. Reeder and H. Rowley, “Time-resolved interventional cardiac C-arm cone-beam CT: an application of the PICCS algorithm”, IEEE Trans. Med. Img., vol. 31, no. 4, pp. 907-923, 2012.

[6] P.J. La Rivière, J. Bian, P.A. Vargas, “Penalized-likelihood sinogram restoration for computed tomography”, IEEE Trans Med Imaging, vol. 25, pp. 1022-1036, 2006.

[7] J. Wang, T. Li, H. Lu, Z. Liang, “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography”, IEEE Trans Med Imaging, vol. 25, pp. 1272- 1283, 2006.

[8] H.A. Ergodan and J.A. Fessler, “Monotonic Algorithms for Transmission Tomography”, IEEE Trans Med Imaging, vol. 18, no. 9, pp. 801-814, 1999.

[9] K. Miles, P. Dawson, M. Blomley, Functional Computed Tomography, ISIS Medical Media, 1997.

[10] T-Y. Lee, “Functional CT: physiological models”, Trends in Biotechnology, vol. 20, no. 8, pp. S3-S10, 2002.

[11] C.M. Kao, J.T. Yap, J. Mukherjee, M.N. Wernick, “Image reconstruction for dynamic PET based on low-order approximation and restoration of the sinogram”, IEEE Trans Med Imaging, vol. 16, pp. 738-749, 1997.

Dynamic Reconstruction with Statistical Ray Weighting for C-Arm CT Perfusion Imaging

Michael T. Manhart, Andreas Fieselmann, Yu Deuerling-Zheng, Andreas K. Maier and Markus Kowarschik

***Abstract*—Tissue perfusion measurement using C-arm angiography systems is a novel technique with potential high benefit for catheter-guided treatment of stroke in the interventional suite. However, perfusion C-arm CT (PCCT) is challenging: the slow C-arm rotation speed only allows measuring samples of contrast time attenuation curves (TACs) every 5 – 6 s if reconstruction algorithms for static data are used. Furthermore, the peaks of the tissue TACs typically lie in a range of 5 – 30 HU, thus perfusion imaging is very sensitive to noise. Recently we presented a dynamic, iterative reconstruction (DIR) approach to reconstruct TACs described by a weighted sum of linear spline functions with a regularization based on joint bilateral filtering (JBF). In this work we incorporate statistical ray weighting into the algorithm and show how this helps to improve the reconstructed cerebral blood flow (CBF) maps in a simulation study with a realistic dynamic brain phantom. The Pearson correlation of the CBF maps to ground truth maps increases from 0.85 (FDK), 0.87 (FDK with JBF), and 0.90 (DIR with JBF) to 0.92 (DIR with JBF and ray weighting). The results suggest that the statistical ray weighting approach improves the diagnostic accuracy of PCCT based on DIR.**

I. Introduction

Perfusion CT (PCT) is an important imaging modality for diagnosis in case of an ischemic stroke event. Time attenuation curves (TACs) in tissue and vessels are extracted from a time series of brain volumes acquired after a contrast bolus injection. Perfusion parameter maps calculated from TACs, which represent quantities such as cerebral blood flow (CBF), cerebral blood volume (CBV), and mean transit time (MTT), provide information about the extent of the affected tissue. They can be used to identify potentially salvageable ischemic tissue that may be reperfused by catheter-guided stroke therapy procedures such as intra-arterial thrombolysis. For this purpose the patient is transported to an interventional suite equipped with a C-arm angiography system, where perfusion measurement is not yet available. Perfusion measurement using C-arm systems would allow assessing the perfusion parameters directly before, during and after the interventional procedure and thus help to determine treatment success and endpoint [1]. Current C-arm systems typically require about 4 – 5 s to acquire the X-ray projection images needed to reconstruct one volume and a pause of about 1 s between two successive acquisitions, which limits the temporal sampling of the TACs and makes perfusion C-arm CT (PCCT) challenging. Recently we proposed a new dynamic, iterative reconstruction algorithm with a joint bilateral filter (DIR-JBF) [2] to reconstruct TACs from a PCCT acquisition with increased temporal resolution and improved CNR in the brain tissue compared to standard FDK reconstruction. In this work, we additionally introduce a statistical ray weighting to further improve the reconstructed perfusion maps. We investigate the noise statistics of subtracted projections and introduce a penalized weighted least squares (PWLS) formulation extending the DIR-JBF algorithm to the DIR *maximum a-posteriori* (DIR-MAP) algorithm. The new DIR-MAP algorithm is evaluated using a digital brain phantom and compared to the DIR-JBF algorithm, classical FDK reconstruction and FDK reconstruction followed by denoising with JBF (FDK-JBF).

M.T. Manhart, A. Fieselmann and A.K. Maier are with Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany. Y. Deuerling-Zheng and M. Kowarschik are with Siemens AG, Healthcare Sector, Angiography & Interventional X-Ray Systems, Forchheim, Germany.

Email: michael.manhart@cs.fau.de

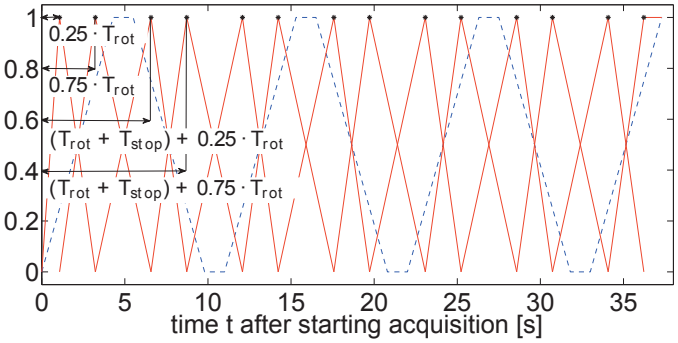


Figure 1. Basis functions for linear interpolation (red, solid) and relative angular C-arm position (blue, dashed).

II. Algorithm

A. Acquisition Protocol

This section describes the C-arm perfusion acquisition protocol used for the simulation study. The parameters are taken from an acquisition protocol available in state-of-the-art C-arm systems (Artis zee, Siemens Healthcare, Germany). Since currently available C-arm systems are not capable of continuous, uni-directional rotations, the C-arm is rotated in a bi-directional manner in forward and backward direction during a perfusion scan. At first one C-arm rotation in forward and one in backward direction acquires mask projection data $\mathbf{p}^{\mathrm{M}}$ (M: mask) with the static anatomical structures. In each rotation 248 projections covering an angular range of 197.6° are acquired. After contrast agent injection the C-arm is rotated $N_{\mathrm{rot}} = 7$ times in bi-directional manner as shown in Figure 1 and acquires the projections $\mathbf{p}^{\mathrm{B}}$ (B: bolus) following the contrast bolus flow. Each rotation takes $T_{\mathrm{rot}} = 4.3\,\mathrm{s}$ with a pause of $T_{\mathrm{stop}} = 1.2\,\mathrm{s}$ between any two of them. After logarithmic

pre-processing the pure contrast-enhanced projection data $\mathbf{p}$ is computed by subtracting the mask projections $\mathbf{p}^{\mathrm{M}}$ from the bolus projections $\mathbf{p}^{\mathrm{B}}$.

B. Dynamic Iterative Reconstruction Algorithm

The DIR-JBF algorithm [2] represents the basis method used to reconstruct the contrast time attenuation curves (TACs) inside the volume of interest (VOI) from the dynamic projection data denoted by vector $\mathbf{p} \in \mathcal{R}^N$, which comprehends all rays during the contrast-enhanced acquisition after mask subtraction. Each TAC inside the VOI is described by a weighted sum of asymmetric linear spline functions. The knots of the splines are placed at $0.25 \cdot T_{\mathrm{rot}}$ and $0.75 \cdot T_{\mathrm{rot}}$ of each C-arm rotation (Figure 1). Consider the 4D volume denoted by vector $\mathbf{x}$ describing the TACs sampled at the acquisition time point of each acquired contrast-enhanced C-arm projection. We introduce the matrix $\mathbf{B}$, which interpolates $\mathbf{x}$ from the weights of all basis functions denoted by vector $\mathbf{w}$ such that $\mathbf{x} = \mathbf{B}\mathbf{w}$. The system matrix $\mathbf{A}$ describes the dynamic forward projection $\mathbf{p} \approx \mathbf{A}\mathbf{x}$. The DIR-JBF algorithm reconstructs the basis weights $\mathbf{w}$ from $\mathbf{p}$ by minimizing the least-squares distance between the measured projection data $\mathbf{p}$ and the forward projected estimated 4D volume:

$$\tilde{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2} \left\| \mathbf{A}\mathbf{B}\mathbf{w} - \mathbf{p} \right\|_2^2 . \tag{1}$$

This large scale problem can be solved as described in [3] by using a gradient-based iterative procedure based on the Landweber scheme:

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \beta \cdot \mathbf{B}^T \mathbf{A}^T \left(\mathbf{p} - \mathbf{A}\mathbf{B}\mathbf{w}^k \right) . \tag{2}$$

The parameter β controls the step size of the parameter update in each iteration, $\mathbf{AB}$ describes a linear interpolation followed by forward projection and $\mathbf{B}^T\mathbf{A}^T$ represents a weighted backprojection of the error image onto the basis weights. As described in [2] the gradient update step is using a vessel-masked backprojection, where rays intersecting with high contrast vessel structures are only backprojected onto voxels belonging to the vessel structures. The weight vector $\mathbf{w}$ is initialized from FDK reconstructions of the C-arm rotations. Furthermore JBF is used for regularization. JBF is an adapted version of the bilateral filter, where the range similarity image is computed using a guidance image. In the DIR-JBF algorithm the temporal maximum intensity projections of the reconstructed TACs are used as guidance image. We modify the JBF regularization compared to [2]: after FDK initialization $N_{\mathrm{JBF}} = 3$ JBF iterations are applied. During the DIR, JBF is applied once after every three gradient update steps. To show the benefits of the dynamic iterative reconstruction algorithm compared to pure FDK reconstruction followed by JBF the result of the initialization is included in the evaluation. Applying only the initialization step is denoted as the FDK-JBF algorithm.

C. Statistical Ray Weighting

In this section we discuss how we model the noise in the subtracted projection data $\mathbf{p}$ and include a statistical noise model into the DIR-JBF algorithm to extend it to the DIR-MAP algorithm. The number of photons measured at the C-arm detector is considered as a Poisson random process to simulate quantum noise. The number of photons reaching a detector pixel is related to the line integral p by $i = i^{\mathrm{S}} \exp(-p)$, where i^{S} denotes the number of photons emitted at the source. Incorporating quantum noise, the number of actually measured photons is a Poisson random process: $\hat{i} \sim \mathcal{P}(\mu_{\mathrm{P}} = i)$ (P: photons). Since we do tomographic brain imaging we assume a large number of counts; i.e. $i > 1000$. For such large counts a Gaussian process is an excellent approximation of the Poisson process:

$$\hat{i} \sim \mathcal{N}\left(\mu_{\mathrm{P}} = i; \sigma_{\mathrm{P}} = \sqrt{i} \right) .$$

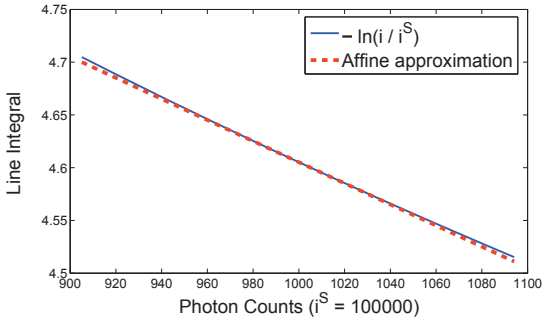


Figure 2. First order Taylor approximation of logarithmic pre-processing around a photon count of $i = 1000$ in the $3\,\sigma_{\mathrm{P}}$ interval of the corresponding Gaussian random process $\hat{i}$ (99.7 % of measurements of 1000 photons will be inside the interval $[1000 - 3\,\sigma_{\mathrm{P}} \quad 1000 + 3\,\sigma_{\mathrm{P}}]$; $\sigma_{\mathrm{P}} = \sqrt{1000}$).

The measured line integrals are also random variables and related to the photon counts by $\hat{p}\left(\hat{i}\right) = -\ln\left(\hat{i}/i^{\mathrm{S}}\right)$. As discussed in [4] for large photon counts we can simplify the logarithmic processing by a first order Taylor series development of $\hat{p}\left(\hat{i}\right)$ around $\mu_{\mathrm{P}} = i$:

$$\begin{aligned} \hat{p}\left(\hat{i}\right) &\approx \hat{p}(i) + \hat{p}'(i)\left(\hat{i} - i\right) \\ &= -\ln\frac{i}{i^{\mathrm{S}}} + \frac{i - \hat{i}}{i} . \end{aligned} \tag{3}$$

Figure 2 shows that the first order development is an appropriate approximation inside the $3\,\sigma_{\mathrm{P}}$ interval of $\hat{i}$. From Equation 3 we see that the logarithmic processing $\hat{p}\left(\hat{i}\right)$ is mainly a scaling with $-1/i$ and shifting with $-\ln\left(i/i^{\mathrm{S}}\right) + 1$. Thus the line integral random variable $\hat{p}$ can also be described as a Gaussian process with mean $\mu_{\mathrm{L}} = \mu_{\mathrm{P}}/i - \ln(i/i_{\mathrm{S}}) + 1 = -\ln(i/i_{\mathrm{S}})$ and variance $\sigma_{\mathrm{L}} = \sigma_{\mathrm{P}}/i$ (L: line integrals):

$$\hat{p} \sim \mathcal{N}\left(\mu_{\mathrm{L}} = -\ln(i/i_{\mathrm{S}}); \sigma_{\mathrm{L}} = 1/\sqrt{i} \right) . \tag{4}$$

Modeling the noise in line integral domain by Gaussian random processes allows to describe the subtraction of the mask measurements from the contrast-enhanced measurements as a subtraction of two independent Gaussian random processes. Thus a mask-subtracted contrast measurement $\hat{p}^{\mathrm{S}} = \left(\hat{p}^{\mathrm{B}} - \hat{p}^{\mathrm{M}}\right)$ (S: subtracted) is again a Gaussian process:

$$\hat{p}^{\mathrm{S}} \sim \mathcal{N}\left(\mu_{\mathrm{L}}^{\mathrm{S}} = p^{\mathrm{B}} - p^{\mathrm{M}}; \sigma_{\mathrm{L}}^{\mathrm{S}} = \sqrt{1/i^{\mathrm{B}} + 1/i^{\mathrm{M}}} \right) . \tag{5}$$

Parameter	FDK	FDK-JBF	DIR-JBF	DIR-MAP
σ_{K}	1.25	0.25	0.25	0.25
σ_{D}		1.5 mm	1.5 mm	1.5 mm
σ_{R0}		0.001	0.001	0.001
σ_{R}		$1.25 \cdot 10^{-4}$	$1.25 \cdot 10^{-4}$	$1.25 \cdot 10^{-4}$
β			3	12
N_{JBF}		3	3	3
N_{DIR}			12	12

σ_{K}: smoothness of FDK filter kernel for initial reconstruction,
σ_{D}: spatial bandwidth of JBF, σ_{R0}: range bandwidth of initial JBF,
σ_{R}: range bandwidth of JBF, β: DIR update step size,
N_{JBF}: number of initial JBF iterations,
N_{DIR}: number of DIR-JBF/DIR-MAP iterations

Table I
PARAMETERS OF ALGORITHMS.

The maximum likelihood (ML) estimation of the weights $\mathbf{w}$ from projection data with Gaussian noise is provided by the corresponding log-likelihood function [5], which combines Equation 1 with the diagonal weighting matrix $\mathbf{D}$ to the squared Mahalanobis distance $\mathcal{D}(\mathbf{w})$:

$$\mathcal{D}(\mathbf{w}) = \frac{1}{2}(\mathbf{ABw} - \mathbf{p})^T \mathbf{D}(\mathbf{ABw} - \mathbf{p}), \qquad (6)$$

where $\mathbf{D} = \mathrm{diag}\left\{1/\left(\sigma^{\mathrm{S}}_{\mathrm{L},1}\right)^2, \ldots, 1/\left(\sigma^{\mathrm{S}}_{\mathrm{L},N}\right)^2\right\}$, with N defined in subsection II-B. Note that in the case of non-subtracted projections the ML estimation would be the same as in well-known statistical reconstruction algorithms [6]. However, our derivation also allows to describe the noise in subtracted projections.

Elad [7] showed that the bilateral filter is related to Bayesian noise removal. Thus we can combine the ML estimate of the weights with a JBF penalty function $\mathcal{R}^{\mathrm{JBF}}(\mathbf{w})$ resulting in the *maximum a-posteriori* (MAP) estimate, which can be formulated as a PWLS problem :

$$\tilde{\mathbf{w}} = \underset{\mathbf{w}}{\mathrm{argmin}}\, \mathcal{D}(\mathbf{w}) + \lambda \mathcal{R}^{\mathrm{JBF}}(\mathbf{w}). \qquad (7)$$

The weights are then updated by the following gradient-based iterative procedure:

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \beta \cdot \mathbf{B}^T \mathbf{A}^T \mathbf{D}\left(\mathbf{p} - \mathbf{ABw}^k\right). \qquad (8)$$

Analogous to the DIR-JBF algorithm, the weights are initialized from FDK reconstructions and filtered $N_{\mathrm{JBF}} = 3$ times after FDK initialization and once after every three gradient update steps. Also vessel masking is applied in the backprojection step. In this work we use the approximation $\mathbf{D} = \mathrm{diag}\left\{\exp\left(-p^{\mathrm{M}}_1\right), \ldots, \exp\left(-p^{\mathrm{M}}_N\right)\right\}$. The contrast attenuation is very small compared to the attenuation of the anatomic structures and thus $\sigma^{\mathrm{S}}_{\mathrm{L},i} \approx \sqrt{2/i^{\mathrm{M}}_i} = \sqrt{2/\left(i^{\mathrm{S}} \exp\left(-p^{\mathrm{M}}_i\right)\right)}$, which results in the above weighting matrix after omitting the constants.

III. MATERIALS & METHODS

We evaluate the different approaches using the realistic digital brain perfusion phantom, which was originally described in [8] and extended for C-arm perfusion imaging in [9]. The phantom is based on segmentation of a human MR brain scan and simulates TACs inside a stroke-affected brain. The phantom is available online [10]. Different regions with reduced blood flow and volume were annotated in the brain likewise as in [9]. The dynamic C-arm projection data was created by forward projecting the 4D phantom according to the acquisition protocol. Poisson-distributed noise was added to the projections assuming an unattenuated X-ray density of $2.1 \cdot 10^5$ photons per mm^2 at the detector. The CBF maps were calculated from the reconstructed TACs using a standard deconvolution-based approach [11]. In this study we compared simple FDK reconstruction with the FDK-JBF, DIR-JBF, and DIR-MAP approaches with the parameters shown in Table I. We computed the RMSE and Pearson correlation (PC) of the resulting CBF maps to reference maps created from ground truth data using an automated ROI analysis [1]. Each of the 18 slices with stroke annotation of the CBF volume was divided into quadratic ROIs of size 8×8 mm^2 and the mean of each ROI was considered as a measurement for RMSE and PC computation. ROIs including voxels outside the brain or vessels were ignored. The slow C-arm rotation speed also leads to artifacts around arteries with high contrast dynamics if FBP-type reconstruction algorithms are used [12]. A qualitative comparison of DIR to FDK reconstruction with respect to such artifacts was performed.

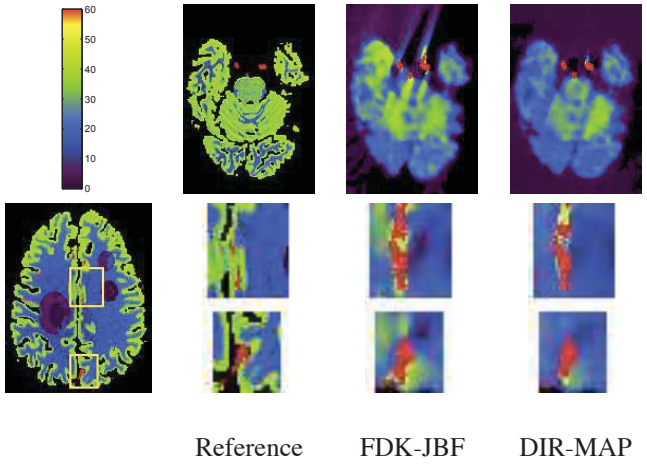


Figure 3. Artifacts in CBF maps (units: ml/100 g/min) comparing DIR-MAP and FDK-JBF reconstructions to the reference for two different slices. In the second row, zoomed views of the CBF maps shown in Figure 4 are provided as indicated by the rectangular regions in the lower left image.

	FDK	FDK-JBF	DIR-JBF	DIR-MAP
RMSE	8.4	4.6	6.2	3.7
PC (n =1815)	0.85	0.87	0.90	0.92

Table II
QUANTITATIVE RESULTS OF CBF MAPS FROM DIGITAL BRAIN PERFUSION PHANTOM DATA RECONSTRUCTED WITH DIFFERENT APPROACHES (RMSE: ROOT MEAN SQUARE ERROR IN [ML/100 ML/MIN], PC: PEARSON CORRELATION TO REFERENCE MAPS WITH SAMPLE SIZE N).

IV. RESULTS

Figure 4 shows the resulting CBF maps reconstructed with the different approaches. The quantitative results are shown in Table II. Figure 3 shows the artifacts around high contrast

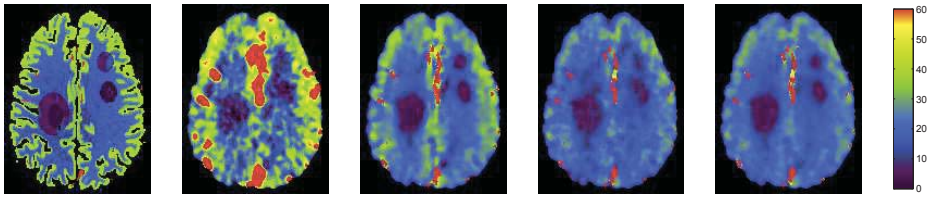


Figure 4. CBF maps (units: ml/100 ml/min) from digital brain perfusion phantom data reconstructed with different approaches.

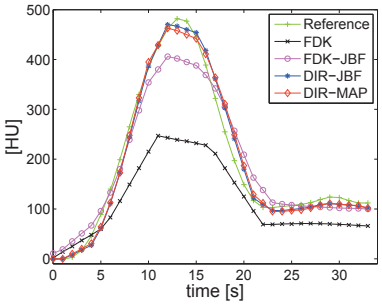


Figure 5. AIFs reconstructed with different approaches compared to ground truth AIF.

vessels in detail for the evaluated algorithms. Figure 5 shows the resulting arterial input functions (AIFs) of the discussed approaches. The reconstructions were performed on a laptop computer with an Intel i7 M 620 2×2.72 GHz CPU, 8 GB RAM, and an Nvidia Quadro FX 880M graphics chip set. The reconstruction of a typical 4D volume of size $256 \times 256 \times 86$ voxels and 14 spline weights per voxel took about 25 min using the DIR-MAP approach and about 1.5 min using the FDK approach, where the projection pre-processing and perfusion parameter computation is not included.

V. Discussion & Conclusions

The CBF maps in Figures 4 and 3 and the AIFs in Figure 5 show how the new DIR-MAP algorithm helps to improve the reconstructed blood flow maps and AIFs in comparison to the other evaluated techniques. The FDK maps have clearly the poorest quality: they are very noisy and the vessels are blurred into the soft tissue due to the very smooth reconstruction kernel. This also leads to a severe underestimation of the AIF and an overestimation of the perfusion values. Furthermore the stroke-affected areas are not well separated from the healthy tissue. The edge-preserving filter in the FDK-JBF reconstruction provides a highly improved noise level in the tissue without blurring the vessels. However, due to low C-arm rotation speed artifacts around high contrast vessels are visible and the AIF is still considerably underestimated. The DIR-JBF algorithm keenly reduces the FBP artifacts around the vessels by including the contrast dynamics into an iterative reconstruction approach and the temporal resolution of the reconstructed AIF is perceptibly improved. However, the resulting perfusion maps look more noisy. By including a statistical noise model, the new DIR-MAP compensates for this drawback. The DIR-MAP technique provides improved results with low noise level, reduced artifacts, improved AIF reconstruction, and physically correct perfusion values. This corresponds to the quantitative results in Table II. The DIR-MAP algorithm yields the best results of all algorithms. Comparing DIR-MAP to standard FDK reconstruction, the RMSE is reduced from 8.4 ml/100 ml/min to 3.7 ml/100 ml/min and PC is increased from 0.85 for standard FDK reconstruction to 0.92 on a sample size of $n = 1815$ ROIs.

In this work we extended our DIR-JBF algorithm [2] by including a statistical ray weighting to the DIR-MAP algorithm. We showed that the DIR-based algorithms help to increase the temporal resolution of the reconstructed TACs and provide an improved estimation of the AIF compared to FDK-type approaches. Furthermore the artifacts around vessels with high contrast are keenly reduced. However, we also found that the recently presented DIR-JBF algorithm produces more noisy perfusion maps than the FDK-JBF technique. By introducing a statistical ray weighting, we can compensate for this drawback.

References

[1] A. Fieselmann, A. Ganguly, Y. Deuerling-Zheng, M. Zellerhoff, C. Rohkohl, J. Boese, J. Hornegger, and R. Fahrig, "Interventional 4-D C-arm CT perfusion imaging using interleaved scanning and partial reconstruction interpolation," *IEEE Trans Med Imaging*, vol. 31, no. 4, pp. 892 – 906, 2012.

[2] M. Manhart, M. Kowarschik, A. Fieselmann, Y. Deuerling-Zheng, and J. Hornegger, "Fast dynamic reconstruction algorithm with joint bilateral filtering for perfusion C-arm CT," in *IEEE NSS MIC Conference Record*, Anaheim, USA, 2012, pp. 2304–2311.

[3] C. Neukirchen, M. Giordano, and S. Wiesner, "An iterative method for tomographic x-ray perfusion estimation in a decomposition model-based approach," *Medical Physics*, vol. 37, no. 12, pp. 6125–6141, 2010.

[4] J. Hsieh, "Adaptive filtering approach to the streaking artifact reduction due to X-ray photon starvation," *Medical Physics*, vol. 25, pp. 2139–2147, 1998.

[5] G. L. Zeng, *Medical Image Reconstruction*. Springer, 2009.

[6] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, p. 1993, 534-548.

[7] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans Img Proc*, vol. 11, pp. 1141–1151, 2002.

[8] A. J. Riordan, M. Prokop, M. A. Viergever, J. W. Dankbaar, E. J. Smit, and H. W. A. M. de Jong, "Validation of CT brain perfusion methods using a realistic dynamic head phantom," *Medical Physics*, vol. 38, no. 6, pp. 3212–3221, 2011.

[9] M. Manhart, A. Fieselmann, and Y. Deuerling-Zheng, "Evaluation of a tight frame reconstruction algorithm for perfusion C-arm CT using a realistic dynamic brain phantom," in *Proc. 2nd CT Meeting*, Salt Lake City, USA, 2012, pp. 123–126.

[10] Digital brain perfusion phantom. http://www5.cs.fau.de/data/.

[11] A. Fieselmann, M. Kowarschik, A. Ganguly, J. Hornegger, and R. Fahrig, "Deconvolution-based CT and MR brain perfusion measurement: Theoretical model revisited and practical implementation details," International Journal of Biomedical Imaging, 2011, article ID 467563.

[12] A. Fieselmann, F. Dennerlein, Y. Deuerling-Zheng, J. Boese, R. Fahrig, and J. Hornegger, "A model for filtered backprojection reconstruction artifacts due to time-varying attenuation values in perfusion C-arm CT," *Physics in Medicine and Biology*, vol. 56, no. 12, pp. 3701–3717, 2011.

Systematic Performance Optimization of Cone-Beam Back-Projection on the Kepler Architecture

Timo Zinßer and Benjamin Keck

***Abstract*—Filtered back-projection algorithms are widely used for the reconstruction of volumetric data from cone-beam projections in interventional C-arm computed tomography. Furthermore, general-purpose GPUs have become a popular tool for accelerating the reconstruction during time-critical clinical procedures. In this work, we focus on the systematic performance optimization of cone-beam back-projection on the latest architecture of CUDA-enabled GPUs. Our optimization approach is based on the identification of the major performance bottleneck through the analysis of specifically modified kernels.**

Our main contribution is a smart restructuring of the back-projection algorithm that facilitates the simultaneous processing of a large number of projections and improves the hit rate of the texture cache at the same time. We use the well-known RabbitCT benchmark to demonstrate the outstanding performance of our implementation on a single Kepler-based GeForce GTX 680 GPU. Our implementation performs the back-projection of 496 input projections onto a cubic 512^3 volume in less than one second, which is three times as fast as the best competing implementation. Our back-projection implementation is also able to reconstruct a cubic 1024^3 volume in about six seconds, which is six times as fast as the best competing implementation known to us.

***Index Terms*—computed tomography, CUDA, FDK, GPGPU**

I. Introduction

There are cone-beam back-projection implementations for a wide range of hardware platforms, including the Cell broadband engine [1], [2], multi-core CPUs [3], [4], and general-purpose GPUs [5]–[7]. Performance comparisons of several implementations have been provided in [1]–[3]. However, the use of varying data sets and diverse reconstruction parameters precludes a meaningful comparison of the existing implementations. This problem has been tackled with the creation of the RabbitCT platform [8], which provides a standardized framework for comparing both the accuracy and the performance of cone-beam back-projection algorithms.

According to [9], filtered back-projection was the first non-graphics compute application to be successfully accelerated on a dedicated GPU. In the meantime, GPUs have evolved into programmable many-core processors, and development platforms like the CUDA framework [10] have simplified the implementation of GPU-accelerated algorithms considerably. Okitsu *et al.* present a comprehensive overview of techniques for the efficient implementation of cone-beam back-projection on CUDA-enabled GPUs in [6]. Their most important contribution is the substantial reduction of memory accesses by processing several projections at a time. Papenhausen *et al.* describe a back-projection implementation that is optimized for Fermi-based GPUs with CUDA support in [7].

As we use the RabbitCT benchmark to evaluate our work, we shortly describe the corresponding data set in Section II. We then discuss several important aspects of GPU computing in Section III. Our approach for systematic performance optimization, as well as the resulting optimized implementation, are presented in Sections IV and V. We analyze the performance of our cone-beam back-projection implementation in Section VI. Finally, Section VII concludes this work.

T. Zinßer and B. Keck are with Siemens AG, Healthcare Sector, Imaging & IT Division, P.O. Box 1266, D-91294 Forchheim, Germany. Corresponding author: Timo Zinßer, E-mail: timo.zinsser@siemens.com.

II. Problem Description

The input data set of the RabbitCT benchmark consists of $N = 496$ projections with a width of $S_u = 1248$ pixels and a height of $S_v = 960$ pixels. For every projection n, the data set also contains a projection matrix $\boldsymbol{P}_n \in \mathbb{R}^{3\times4}$, which describes the transformation from the world coordinate system to the detector coordinate system. An illustration of the acquisition geometry can be found in [4]. Basically, the detector rotates on a circular short-scan trajectory around the z-axis of the world coordinate system, which is also the z-axis of the reconstructed volume. This axis is roughly perpendicular to the u-axis of all projections, as well as roughly parallel to their v-axis.

The task of a back-projection algorithm in the RabbitCT benchmark is the reconstruction of a cubic volume with a side length of 256 mm. The benchmark defines three problem sizes, which correspond to volumes with a side length of 256, 512, or 1024 voxels, respectively. During the kernel optimization in Section IV, we focus on the 512^3 volume. As every projection has to be back-projected onto every voxel, the reconstruction of the 512^3 volume requires approximately 66.6×10^9 voxel updates. This value is important for computing a common alternative performance measure, the giga-(voxel)-updates per second (GUPS). Please take note that we strictly differentiate between decimal prefixes (1 gigabyte = 1 GB = 10^9 bytes) and binary prefixes (1 gibibyte = 1 GiB = 2^{30} bytes) to prevent unnecessary ambiguities in this work.

III. GPU Computing

In our experience, the full performance potential of any new hardware platform can only be realized by specifically optimizing the implementation of the deployed algorithms. The main hardware platform of this work is the Kepler-based GeForce GTX 680 GPU, which is compared to its predecessors

TABLE I
RECENT GENERATIONS OF HIGH-END GPUS WITH CUDA SUPPORT

GPU	GTX 280	GTX 480	GTX 580	GTX 680
Architecture	Tesla	Fermi	Fermi	Kepler
Performance [GFLOPS]	622.1	1345.0	1581.1	3090.4
Texture fillrate [GT/s]	48.2	42.0	49.4	128.8
Bandwidth [GB/s]	141.7	177.4	192.4	192.3

in Table I. Evidently, the arithmetic throughput of these GPUs has increased with every generation. It is important to note that the specified peak performance is only achieved for fused multiply-add operations. For other operations, the performance is reduced at least by a factor of two. The texture fillrate has remained almost constant for several generations, but it has more than doubled for the GeForce GTX 680. In contrast to this, the memory bandwidth has all but stagnated in the latest generation. As a consequence, the ratio between the texture fillrate and the memory bandwidth has also more than doubled. In order to save memory bandwidth and fully utilize its texture units, the GeForce GTX 680 buffers texture data in the read-only caches of its streaming multiprocessors as well as in its unified L2 cache. For convenience, we refer to the combination of these caches as texture cache.

We use the CUDA framework for implementing GPU-based algorithms. An overview of basic concepts of GPU computing can be found in [9]. For advanced topics, we recommend the documentation of the CUDA framework itself [10]. One important aspect of optimization is the identification of the current performance bottleneck. Typically, the performance of an algorithm is either limited by the instruction throughput, the memory bandwidth, or the texture fillrate. In contrast to this, latencies are usually not a problem, as long as the algorithm exposes enough thread-level or instruction-level parallelism.

Compared to a sequential CPU algorithm, the execution order of a highly parallel GPU algorithm is more complicated and less deterministic. The following overview identifies various levels of temporal cohesion on a Kepler-based GPU:

- The 32 threads of one warp run in lockstep. They can communicate via very efficient shuffle instructions.
- One or more warps form a thread block. The threads in one thread block can communicate via shared memory. These threads can also be synchronized explicitly.
- One or more threads block are executed on a streaming multiprocessor. These thread blocks share the L1 cache and the read-only cache of the multiprocessor. The ratio between the actual number of threads on a multiprocessor and the theoretical maximum is called occupancy.
- In general, only a subset of all thread blocks fits onto the available multiprocessors at the same time. This subset is called a wave. The thread blocks of a wave share the unified L2 cache of the GPU.

Iterative loops inside a kernel exhibit less temporal cohesion than the threads of a warp, and may have less temporal cohesion than the threads of a thread block, if these threads are synchronized after every iteration. The execution order of instructions in different thread blocks is not defined by the CUDA programming model.

```
compute position of first voxel
for I input projections do
    compute homogeneous detector coordinates q[i] of first voxel
end
for K consecutive voxels along the z-axis do
    zero-initialize sum s of weighted back-projected values
    for I input projections do
        dehomogenize detector coordinates q[i]
        compute back-projected value by texture fetching
        update sum s of weighted back-projected values
        update homogeneous detector coordinates q[i]
    end
    update volume at current voxel with computed sum s
    (optionally) synchronize threads in thread block
end
```

Fig. 1. Pseudocode for cone-beam back-projection kernel A

IV. KERNEL OPTIMIZATION

The cone-beam back-projection kernel presented in Fig. 1 constitutes the starting point of our performance optimization. The structure of this kernel is based on the structure of what is referred to as the fully optimized configuration in [7]. In our kernel, one thread updates K voxels with the weighted back-projected values of I projections. Every thread block consists of $B_x \times B_y$ threads. The voxels updated by a single thread block constitute a rectangular tile in the respective consecutive xy-slices of the volume. A volume with $S_x \times S_y \times S_z$ voxels is processed by a grid of $(S_x/B_x) \times (S_y/B_y) \times (S_z/K)$ thread blocks. We use layered textures to simplify the texture fetching for several projections. We also store the projection matrices in constant memory, which is backed by the read-only cache. Finally, we specify the number of projections as a template parameter, which allows the compiler to automatically unroll the corresponding loops of kernel A.

In this section, we ignore all data transfers involving the host and focus on the computation times of the kernels for the 512^3 volume. In order to identify the performance bottleneck of a configuration, we measure three additional computation times. In the first step, we reduce the voxel size from 0.5 mm to 10^{-6} mm. As a result, all computed detector coordinates are virtually identical, and the hit rate of the texture cache rises to almost one hundred percent. In the second step, we disable the texture fetching completely. In the third step, we also turn off the volume update, which removes the memory accesses and leaves only the arithmetic and control flow instructions. It is vital that these modifications do not allow the compiler to eliminate more code than intended. As these modifications also tend to reduce the register count, we allocate a suitable amount of shared memory to retain the occupancy of the original kernel. Using the letters I(nstruction), M(emory), and T(exture), we label the corresponding additional computation times as I|M|T, I|M|−, and I|−|− in Table II.

In our first test, kernel A processes one projection at a time. Each thread updates one voxel in every xy-slice of the volume. The specified tile width $B_x = 32$ ensures that the volume updates are performed by fully coalesced memory transactions. Nevertheless, the first row of Table II clearly shows that the memory transfer takes much longer than the computation of the arithmetic instructions. In addition to that, the computation

TABLE II
PERFORMANCE BOTTLENECK ANALYSIS FOR DIFFERENT KERNEL CONFIGURATIONS

Test	Kernel	Sync	I	K	B_x	B_y	Occupancy	I | – | –	I |M| –	I |M| T	Time	GUPS
1	A	no	1	512	32	8	1.000	1091 ms	4710 ms	4707 ms	9141 ms	7.3
2	A	no	4	512	32	8	0.750	575 ms	1199 ms	1196 ms	7802 ms	8.5
3	A	yes	4	512	32	8	1.000	554 ms	1185 ms	1169 ms	2710 ms	24.6
4	A	no	4	8	32	8	0.750	685 ms	980 ms	1085 ms	1990 ms	33.4
5	B	—	4	8	32	8	0.875	542 ms	989 ms	1179 ms	1527 ms	43.6
6	B	—	8	4	16	16	0.750	528 ms	725 ms	966 ms	1296 ms	51.4
7	B	—	16	4	16	32	0.750	506 ms	550 ms	826 ms	1051 ms	63.4
8	B	—	32	4	16	32	0.750	489 ms	494 ms	756 ms	969 ms	68.7

```
compute position of first voxel
for K consecutive voxels along the z-axis do
    zero-initialize sum s[k] of weighted back-projected values
end
for I input projections do
    compute homogeneous detector coordinates q of first voxel
    for K consecutive voxels along the z-axis do
        dehomogenize detector coordinates q
        compute back-projected value by texture fetching
        update sum s[k] of weighted back-projected values
        update homogeneous detector coordinates q
    end
end
for K consecutive voxels along the z-axis do
    update volume at current voxel with computed sum s[k]
end
```

Fig. 2. Pseudocode for cone-beam back-projection kernel B

time of the kernel is almost doubled by the cache misses of the texture fetching. When we process four projections in one kernel, the memory transfer size is reduced considerably. The compute-only kernel also runs much faster, because the number of integer-based index computations is minimized as well. However, the time penalty induced by the cache misses of the texture fetching remains very high.

In the third test, we activate the optional synchronization as indicated in Fig. 1. This change prevents the divergence of the threads in one thread block with respect to the loop over the voxels along the z-axis. As a result, the texture fetching is accelerated considerably and the computation time is reduced by about 65 percent. The configuration of test 3 results in a total of 16 waves of thread blocks, which iterate through the volume along the z-axis one after another. In test 4, we relocate the large scale movement along the z-axis from the loop inside the kernel to the third dimension of the grid of thread blocks. On the whole, the 1024 generated waves move through the volume along the z-axis only once, which improves the hit rate of the texture cache even more.

In all tests with kernel A, the cache misses of the texture fetching constitute the major performance bottleneck. As the innermost loop of this kernel iterates over different projections for $I > 1$, the corresponding textures continuously contend for the limited amount of cache memory. Furthermore, the memory transfers for the volume update take longer than the computations. This problem could be solved by increasing the number of projections I, but this approach only exacerbates the first problem. We propose to solve both problems by reversing the order of the two nested loops in kernel A. The result of this restructuring is illustrated in Fig. 2. In kernel B, we specify both I and K as template parameters. All iterations of the innermost loop of this kernel access the same texture. While kernel A uses $3I$ registers to store the homogeneous detector coordinates, kernel B requires K registers for buffering the computed volume updates. Consequently, our proposed kernel is able to process a very large number of projections.

In test 5, we replace kernel A with kernel B, but keep all other parameters identical. We clearly observe an improved hit rate of the texture cache. In the following three tests, we increase the number of projections I and tune the other parameters to obtain minimal computation times. For $I = 32$ projections, the instruction throughput is not the bottleneck and the impact of the memory transfer is negligible. The computation time of the I |M| T modification indicates that our proposed kernel reaches more than 68% of the theoretical texture fillrate. The impact of the cache misses of the texture fetching has also been reduced, resulting in a total computation time of less than one second for the 512^3 volume.

V. DATA TRANSFER OPTIMIZATION

For a useful comparison of our GPU-based implementation with other hardware platforms, the data transfers between the host and the GPU have to be taken into consideration. The practically relevant data transfers consist of the upload of the input projections and the download of the reconstructed volume. The reconstruction of the 512^3 volume of the RabbitCT benchmark results in the transfer of 2779 MiB of data, which takes about half a second on our system. In order to hide the additional time required for the described data transfers, we use the ability of our GPU to overlap kernel execution and data transfer. To this end, the CUDA API allows asynchronous kernel launches and provides asynchronous memcopy functions. However, there are no asynchronous API functions for binding textures, which complicates both the memory management and the texture handling in our implementation.

The timeline in Fig. 3 illustrates the data transfers and kernel executions during the reconstruction of the 512^3 volume. As the first data transfer cannot be overlapped with any kernel launch, we start with $I = 8$ projections to keep the size of this data transfer small. We add eight more projections in every subsequent data transfer until we reach the optimal value of $I = 32$ projections. As a second optimization, we divide the reconstructed volume into two parts, which consist

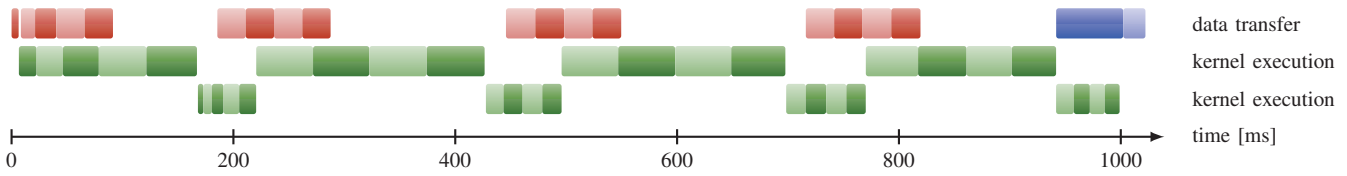


Fig. 3. This figure displays the timeline of the reconstruction of the 512^3 volume. The first line represents the data transfers, which comprise the upload of the projections and the download of the volume. The other two lines illustrate the reconstruction of the first and the second part of the volume, respectively. Please note that the total computation time in this timeline is larger than one second due to profiling overhead.

TABLE III
COMPARISON OF SELECTED CONE-BEAM BACK-PROJECTION IMPLEMENTATIONS LISTED ON THE RABBITCT HOMEPAGE

Volume	Implementation	Type	RMSE	Time	GUPS
512^3	fastrabbitEX [4]	CPU	—	7.45 s	8.94
	RapidRabbit [7]	GPU	—	2.98 s	22.3
	Thumper [this work]	GPU	0.021 HU	0.99 s	67.7
1024^3	fastrabbitEX [4]	CPU	—	43.8 s	12.2
	CERA [-]	GPU	—	36.1 s	14.7
	Thumper [this work]	GPU	0.021 HU	6.04 s	88.2

of 384 and 128 xy-slices, respectively. In combination with the buffering of a certain number of projections on the GPU, this optimization makes it possible to overlap the download of the first part of the volume with the reconstruction of the second part of the volume.

Our implementation also works with the 1024^3 volume of the RabbitCT benchmark. However, this volume has a size of 4096 MiB, which is twice as large as the device memory of our GPU. As there are no data dependencies between different voxels or between different projections, we are free to adapt the high-level data flow of our implementation accordingly. We have chosen to upload the first half of the projections onto the GPU and stream the 1024^3 volume using two buffers with a size of 256 MiB each. We repeat this process for the second half of the projections. Performance measurements for this volume size are provided in the next section.

VI. EXPERIMENTAL RESULTS

Our presented cone-beam back-projection implementation was evaluated on a GeForce GTX 680 GPU with 2048 MiB RAM using version 4.2 of the CUDA framework. In Table III, we compare the obtained results to the best competing implementations listed on the RabbitCT homepage [11]. Our implementation, alias Thumper, has a total computation time of less than one second for the 512^3 volume. This is three times as fast as the best competing implementation RapidRabbit. As both implementations were tested on the same GPU model, the performance difference can be fully attributed to our proposed optimizations. Furthermore, our implementation has a total computation time of about six seconds for the 1024^3 volume. This is six times as fast as the best competing implementation, which uses a slower Fermi-based Tesla C2070 GPU.

The fastest CPU-based implementation in the RabbitCT benchmark is called fastrabbitEX. Although a workstation with 40 CPU cores was used to evaluate fastrabbitEX, our implementation is more than seven times as fast for both volume sizes. Due to an error in the reference volume of the benchmark, we are unable to specify the accuracy of the competing implementations in Table III. When we use the corrected reference volume in floating-point format to assess the accuracy of our implementation, we observe a root mean square error of only 0.021 HU for both volume sizes.

VII. CONCLUSION

Using a systematic performance optimization approach, we identify the bottlenecks of a state-of-the-art cone-beam back-projection kernel. Our main contribution is a restructuring of the kernel that deals with the two most prominent performance bottlenecks all at once. Our implementation is three times as fast as the best competing implementation for the clinically relevant 512^3 volume of the RabbitCT benchmark. Although we focus on the optimization of a cone-beam back-projection kernel for GPUs of the Kepler architecture, our optimization approach is applicable to a wide range of GPU-based algorithms. Furthermore, cursory tests with other CUDA-enabled GPUs have confirmed the portability of our optimizations.

REFERENCES

[1] M. Kachelrieß, M. Knaup, and O. Bockenbach, "Hyperfast parallel-beam and cone-beam backprojection using the Cell general purpose hardware," *Medical Physics*, vol. 34, no. 4, pp. 1474–1486, April 2007.

[2] H. Scherl, M. Kowarschik, H. Hofmann, B. Keck, and J. Hornegger, "Evaluation of State-of-the-Art Hardware Architectures for Fast Cone-Beam CT Reconstruction," *Parallel Computing*, vol. 38, no. 3, pp. 111–124, March 2012.

[3] H. Hofmann, B. Keck, C. Rohkohl, and J. Hornegger, "Comparing performance of many-core CPUs and GPUs for static and motion compensated reconstruction of C-arm CT data," *Medical Physics*, vol. 38, no. 1, pp. 468–473, January 2011.

[4] J. Treibig, G. Hager, H. Hofmann, J. Hornegger, and G. Wellein, "Pushing the limits for medical image reconstruction on recent standard multicore processors," *International Journal of High Performance Computing Applications*, 2012, OnlineFirst.

[5] D. Riabkov, X. Xue, D. Tubbs, and A. Cheryauka, "Accelerated cone-beam backprojection using GPU-CPU hardware," in *Proceedings of the Workshop on High Performance Image Reconstruction*, Lindau, Germany, July 2007, pp. 68–71.

[6] Y. Okitsu, F. Ino, and K. Hagihara, "High-Performance Cone Beam Reconstruction Using CUDA Compatible GPUs," *Parallel Computing*, vol. 36, no. 2-3, pp. 129–141, February 2010.

[7] E. Papenhausen, Z. Zheng, and K. Mueller, "GPU-Accelerated Back-Projection Revisited: Squeezing Performance by Careful Tuning," in *Proceedings of the Workshop on High Performance Image Reconstruction*, Potsdam, Germany, July 2011.

[8] C. Rohkohl, B. Keck, H. Hofmann, and J. Hornegger, "RabbitCT - an open platform for benchmarking 3D cone-beam reconstruction algorithms," *Medical Physics*, vol. 36, no. 9, pp. 3940–3944, 2009.

[9] G. Pratx and L. Xing, "GPU computing in medical physics: A review," *Medical Physics*, vol. 38, no. 5, pp. 2685–2697, May 2011.

[10] "NVIDIA CUDA Toolkit Documentation," 2013. [Online]. Available: http://docs.nvidia.com/cuda/index.html

[11] "Homepage of the RabbitCT project," 2013. [Online]. Available: http://www5.cs.fau.de/research/projects/rabbitct

Fully 3D PET List-Mode reconstruction including an accurate detector modeling on GPU architecture

Awen Autret, Julien Bert, Olivier Strauss and Dimitris Visvikis

***Abstract*—List-mode based image reconstruction is continuously gaining ground in Positron Emission Tomography (PET) imaging. Such reconstruction has many benefits, as the preservation of the temporal sampling of the acquired data that facilitates the study and correction of dynamic processes, as well as maintaining the highest spatial sampling available for a given detector geometry. The centerpiece of the reconstruction process is the projector. This projector computes on-the-fly the contribution of a line-of-response for each voxel of the field-of-view considering detector effects. In this study, we propose a new projector that incorporates detector modeling including geometric and detector scatter effects to improve the reconstructed image accuracy. As the computation burden is a main obstacle to obtain such reconstruction, we implemented the reconstruction including detector corrections on graphic processing units (GPU). Results showed that our projector provides reconstructed images with a high accuracy (low noise, high contrast and resolution).**

***Index Terms*—PET reconstruction, detector modeling, IRIS projector, intrinsic detector response function (IDRF), GPU**

I. Introduction

POSITRON Emission Tomography (PET) is a nuclear medical imaging modality that estimates the 3D image of functional processes in the body. List-mode (LM) based PET reconstruction has several advantages [1] over conventional approaches. Advantages include preservation of the high frequency and high spatial resolution of the acquired data, which in turn facilitate a better handling of dynamic processes as well as allow a finer image spatial resolution.

In 3D PET reconstruction, the system response matrix (SRM), which models the various effects from the imaging system and the patient, plays a key role in both the qualitative and quantitative performances of the reconstructed images. Due to its high complexity and enormous size, this matrix may be decomposed in multiple sub-matrices, each one accounting for a different effect of the detection process [2].

Within the LM context, the sub-matrix modeling the probability density function (PDF) for each line-of-response (LOR) to detect an annihilation from a specific position, can be computed on-the-fly using a projector. The physical reality of this PDF is a 3D volume commonly named volume-of-response (VOR), which has a complex varying shape according to the LOR, its position in the field-of-view (FOV) and the physical effects such as detector scatter.

In classical PET image reconstruction using a projector, the Siddon ray-tracer is used as projector [3]. This basic projector does not handle any PET detector effects correction and consider the VOR as a simple voxelized line. A recent work proposes a projector based on Gaussian functions to model the VOR shape [4]. However, a simple function does not allow an accurate modeling, especially for crystal scatter effects. In this work, we propose a new projector that incorporates detector corrections, including geometric and detector scatter effects, providing the true complex varying shape of the VOR.

Therefore, such LM PET reconstruction suffers from a very high computational time cost incompatible with clinical applications. Recently, graphics processing units (GPU) are becoming the most suitable solution to resolve computational time cost problems, due to their highly parallel architecture. Within this context, a few PET image reconstruction implementations designed for the GPU architecture have been recently proposed [5], [6], tending to reach a run-time compatible with the clinical constraints.

A. Autret, J. Bert and D. Visvikis are with the INSERM UMR1101, LaTIM, CHRU Brest, France (e-mail: awen.autret@telecom-bretagne.eu).

O. Strauss is with the CNRS UMR5506, LIRMM, Université Montpellier 2, France.

II. Materials and methods

A. LM-OSEM reconstruction

List-mode ordered-subsets expectation maximization (LM-OSEM) PET reconstruction algorithm [7] was introduced to accelerate the reconstruction time of list-mode expectation maximization algorithm (LM-EM) [8]. An image update takes the form,

$$\lambda_j^{m,l+1} = \frac{\lambda_j^{m,l}}{\sum_{i=1}^{I} w_{ii} P_{ij}} \sum_{k \in S_l} P_{i_k j} \frac{1}{\sum_{b=1}^{J} P_{i_k b} \lambda_j^{m,l}} \qquad (1)$$

where $\lambda_j^{m,l}$ is the image estimation in the voxel j for the m^{th} iteration and l^{th} list-mode subset, P_{ij} (refers a SRM element) is the probability of an emission from the voxel j to be detected along the LOR i, w_{ii} is a sensitivity matrix element and finally i_k refers to the LOR index of the k^{th} list-mode event stored in the data subset S_l. The image is updated after every subset processed defining a complete iteration and iterations are repeated until convergence. During the reconstruction process one of the major issue is to access the P_{ij} elements, because of the huge size of the SRM. Our work is based on the use of a projector which computes on-the-fly the P_{ij} coefficients related to each VOR i.

B. Detector modeling

In this reconstruction context, the main obstacle is to propose a projector that includes an accurate modeling of the PDF of the LORs (i.e. VOR) according to the PET detector effects.

A LOR is detected by two crystal detectors and the first interaction positions of detected photons within both detectors are closely related to the LOR-PDF. These interaction positions within a crystal define a PDF called intrinsic detector response function (IDRF). Authors in [9] proposed a way to bind the LOR-PDF with the two involved IDRFs, i.e. those provided by the detector pair that recorded the LOR. The relationship between IDRF pair and the LOR-PDF is defined by the following stochastic equation,

$$X = \frac{1}{2}(1-t)Y_1 + \frac{1}{2}(1+t)Y_2 \tag{2}$$

where X is the random variable (RV) of the LOR-PDF, (Y_1, Y_2) are the RVs of both IDRFs linking the LOR, and t a RV uniformly distributed between $[-1, 1]$.

Assuming the IDRFs and the PDF of t are known, it is possible to estimate the PDF of the RV X (i.e. the VOR). One possibility, proposed by [10], consists in using a random sampling approach that can be seen as a Monte Carlo estimator. From random samples of Y_1, Y_2 and t, samples of X can be calculated by applying the relationship (2). The resulting samples are then used to estimate the PDF of X.

Starting from this idea we proposed a new iterative random IDRF sampling (IRIS) projector which includes an accurate model of the IDRF, considering the detector geometry and the detector scattering effects. The rendering process of a VOR lies in iteratively generating two random points, one from each two IDRF models related to the LOR, and rendering the thin voxelized line connecting these two points. This step is repeated, accumulating the rendered thin lines, until the VOR estimation is consistent. This multiple-lines approach allows rendering a VOR varying naturally inside the field-of-view.

The IDRF model used by the IRIS projector is spatially split in two parts. One inside the given crystal detector is related to the geometric model of the crystal including its rectangular shape, angle and position in the space and intra-crystal penetration. A second one, in the neighbor crystals, takes into account the inter-crystal scattering effects and the inter-crystal penetration. The value P_{in} defines the probability of the first scattering position to be in the first part of the model and $(1 - P_{in})$ is the probability to be in the second part. The first part is modeled with a 3D function within the given crystal and composed of a 2D uniform distribution for the front plan of the crystal and a decreasing exponential distribution with a decreasing factor λ_{in} in the depth direction. The second part is modeled using a 3D function with a prohibited region inside the involved crystal. The model is composed of a decreasing exponential function along the crystal depth direction with λ_{out} as decreasing factor and, in the orthogonal plan, a decreasing exponential function with a decreasing factor $\lambda_{\perp out}$ and a circular symmetry with O_{sym} as varying center of symmetry. Values of these five parameters depend on the LOR and, instead of storing every IDRF model required for the reconstruction, we parameterized this model as a function of the incidence angle α between the LOR and the crystal detector. This parameterization allows computing on-the-fly the IDRF model for any LOR. Due to the angular

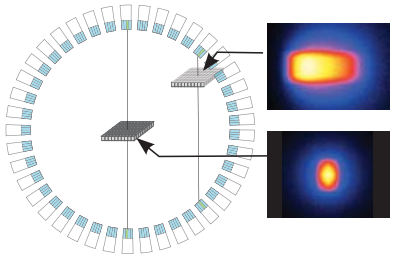


Fig. 1. The two studied VOR sections, a centered one and a shifted one. Images represent sections of each studied VOR recovered from Monte Carlo simulation.

parameterization, values of the model are parameterized using trigonometric functions as described below,

$$\begin{aligned} P_{in} &= a_1 \cos(b_1 \alpha) \\ \lambda_{in} &= \frac{1}{a_2 \cos(\alpha)} \\ \lambda_{out} &= \frac{1}{a_3 \cos(\alpha)} \\ \lambda_{\perp out} &= Constant \\ O_{sym} &= a_4 \sin(\alpha) \end{aligned} \tag{3}$$

where a_1, b_1, a_2, a_3, $Constant$ and a_4 are the global parameters of the parameterization of the IDRF model. It has the advantage of fully model the detector and varies depending on the angle of incidence of the LOR in the crystal.

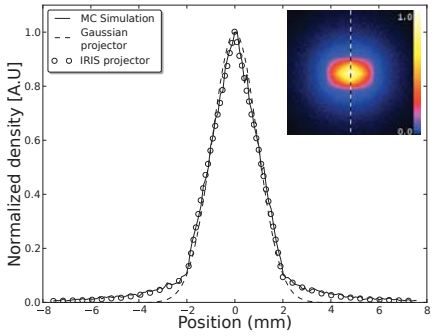
C. GPU Implementation

A GPU is a massively parallel processor that is composed of thousands of threads, each one representing a data processing unit. All threads will execute the same program code, called kernel, in parallel on the different streaming processors, representing individual processing units. We choose a strategy that consists of executing one thread per LOR. This implementation is based on NVIDIA GPUs, using its parallel computing platform CUDA.

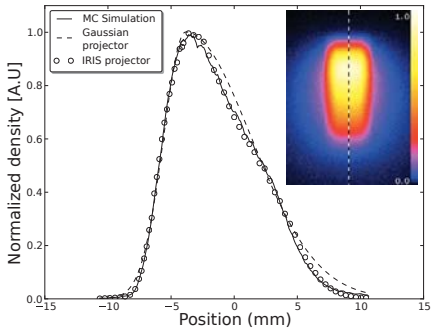
D. Evaluation Study

This study was based on a Philips GEMINI PET scanner model [11] and performed with the Monte Carlo simulation (MCS) GATE platform [12]. Measurements provided by MCS are considered as gold standard. The IDRF model including its parameterization and the Gaussian model were estimated from measurements provided by MCS. Measured PSF FWHM varied from 2.4mm in the center of the FOV to 8.4mm in the edge.

We first validated the capability of the IRIS projector to build realistic and accurate VORs. Due to the high computation cost to perform MCS, our study was based only on sections of two vertical and parallel LORs. The first one was centered on the FOV and the second one was radially shifted by 216mm. One specific section per VOR was studied, namely at the LOR center for the first LOR and at a tangential position of 156mm from the LOR center for the second one, as illustrated in Fig. 1. These two VORs were rendered with the



(a)



(b)

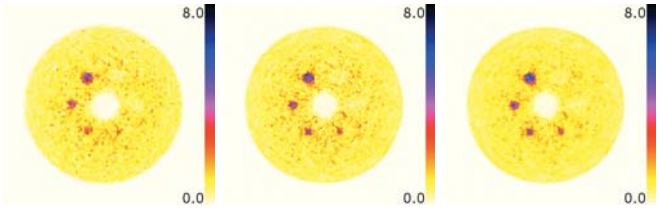
Fig. 2. Profiles through the sections from the (a) central LOR and (b) radially shifted LOR. Sections were estimated from the MCS and both Gaussian and IRIS projectors. To the upper right, the vertical lines on VOR sections recovered from MCS indicate the location of the profiles.

IRIS and Gaussian projectors for a resolution of $0.1^3 mm^3$. For the IRIS projector the rendering was performed using 50.10^6 random lines to minimize the noise. The VOR sections were compared to the respective measurements provided by MCS.

Finally, we evaluated the IRIS projector against the Siddon and Gaussian projector within a reconstruction context. The first studied phantom was the NEMA IEC, which is composed of a warm water cylinder containing four hot spheres of 10, 13, 17 and 22mm diameters and two cold spheres of 28 and 37mm diameters. The list-mode dataset was obtained from MCS and 12×10^6 true unscattered coincidences were used. The second phantom was a miniature Derenzo type composed of a plastic cylinder of 45mm diameter and 16mm height with six sections of hot activity rods of 1.2, 1.6, 2.4, 3.2, 4.0 and 4.8mm diameters. The dataset provided from the MCS was composed of 2.4×10^6 true unscattered coincidences.

Both phantoms were reconstructed using the LM-OSEM algorithm including normalization and attenuation correction, with 1 subset and 30 iterations for the NEMA IEC NU2-2001 phantom and 3 subsets and 30 iterations for the miniature Derenzo type phantom. Reconstructions were provided using the Gaussian and IRIS projectors on one CPU core and one GPU with voxels of $1^3 mm^3$ for the NEMA IEC phantom and $0.5^3 mm^3$ for the miniature Derenzo type phantom. Each VOR estimated with the IRIS projector were sampled using 100 thin lines.

For the Siddon projector the VOR is considered as a simple voxelized line with a size depending of the crystal detector. The voxel size of the reconstructed images is then constraint



(a) (b) (c)

Fig. 3. Central slice through the reconstructed images of the NEMA phantom using the (a) Siddon projector with voxels of $4^3 mm^3$ (CPU), (b) Gaussian projector with voxels of $1^3 mm^3$ (GPU) and (c) IRIS projector with voxels of $1^3 mm^3$ (GPU).

to the crystal size. Based on the PET scanner model used in this study, reconstructions using the Siddon projector were with a voxel size of $4^3 mm^3$. The Siddon projector was implemented only on one CPU core. Reconstructions were performed with one GPU of a bi-GPU NVIDIA GTX590 (512 cores - 1.23GHz) and an Intel Core i7 CPU (3.4GHz).

From the reconstruction of the NEMA IEC the contrast recovery coefficient (CRC) was evaluated in the smallest (10mm diameter) and the biggest (22mm diameter) hot spheres from the equation above,

$$CRC = \frac{\overline{r_h} - \overline{r_b}}{\overline{r_b}} \tag{4}$$

where, $\overline{r_h}$ is the mean of the voxel values inside the hot sphere, and $\overline{r_b}$ is the mean of the voxel values in the background warm cylinder. The signal-to-noise ratio (SNR) was also evaluated on the homogeneous warm background based on [13] using the equation above,

$$SNR = \frac{\sqrt{2}}{S} \sum_{i}^{S} \frac{a_i}{dsd_i} \tag{5}$$

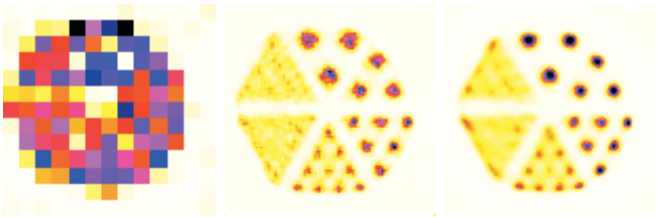
where a_i is the average of all pixels within the background in slice i and S indicates the total number of slices. dsd_i is the standard deviation between the different data as described above,

$$dsd_i = \sqrt{\frac{n \sum_j {d_j}^2 - (\sum_j d_j)^2}{n(n-1)}} \tag{6}$$

where d_j is the difference between the pixel j of two reconstructed images from different datasets and n is the number of pixels in the region-of-interest in the slice i.

III. Results

Profiles comparison between VOR cross sections from the MCS, the Gaussian and IRIS projectors are presented in Fig. 2. Profiles between sections recovered by the Gaussian projector and the MCS are close but not completely identical. Large differences on profile tails indicate that the Gaussian projector does not model properly scattering effects. On the other hand, results between sections obtained by the IRIS projector and the MCS show identical profiles. This suggests that the IRIS projector is able to build accurate VOR including geometric and scatter effects.



(a) (b) (c)

Fig. 4. Central slice through the reconstructed images of the Derenzo phantom using the (a) Siddon projector with voxels of $4^3 mm^3$ (CPU), (b) Gaussian projector with voxels of $0.5^3 mm^3$ (GPU) and (c) IRIS projector with voxels of $0.5^3 mm^3$ (GPU).

Reconstructed images of the NEMA phantom using the Siddon projector with one CPU core, the Gaussian and IRIS projectors on one GPU are shown in Fig. 3. Each projector implemented on both CPU and GPU has reconstructed identical images. Results on the evaluation study are presented on the Table I. The CRCs in the biggest and smallest hot spheres are respectively 3.1 and 1.8 using the Siddon projector, 3.4 and 2.4 using the Gaussian projector (GPU) and 3.6 and 2.4 using the IRIS projector (GPU). The SNRs are 2.9, 2.9 and 3.7 respectively for the Siddon, Gaussian and IRIS projectors. These results show that the IRIS projector gave a better or equivalent CRC with a lower noise on the reconstructed images.

Reconstructed images from the Derenzo phantom using Siddon projector on one CPU core, the Gaussian and IRIS projectors on one GPU are shown in Fig. 4. A visual assessment can be drawn between the methods. From the result given by the Siddon projector, due to the large voxel size and basic modeling, it is not possible to distinguish any rods, even the largest ones of 4.8mm. When the Gaussian and IRIS projectors are used, the rods are noticeable until 2.4mm. The detector correction included in the IRIS projector largely surpasses the Siddon projector in term of spatial resolution. Compared to the Gaussian projector, the main difference with the IRIS projector is the level of noise which seems smaller, improving the rods structure recovering.

The reconstruction times in Table I show that the Gaussian and IRIS projectors using one CPU core are completely incompatible with the clinical time. Including the detector correction raises the projector complexity and dramatically drops the reconstruction time compared to a simple projector like Siddon. However, in the case of the IRIS projector, our GPU implementation improves this run-time and can reach a speedup of 10 with one GPU. Considering both CPU and GPU implementations, the IRIS projector is faster than the Gaussian projector, respectively with a speedup factor of 9 and 1.9 for the CPU and GPU versions.

TABLE I
MEASUREMENTS ON RECONSTRUCTED IMAGES

	Siddon CPU	Gaussian CPU	Gaussian GPU	IRIS CPU	IRIS GPU
CRC_1	3.1	3.4	3.4	3.6	3.6
CRC_2	1.8	2.4	2.4	2.4	2.4
SNR	2.9	2.9	2.9	3.7	3.7
Run time	3.1s	13633s	288s	1527s	151s

Times are given for 10^6 *processed events per iteration.*

IV. CONCLUSION

In this work, we proposed a new projector that provides detector corrections including at the same time the scanner geometry and the detector scattering effects, with a high performance implementation of a 3D LM-OSEM PET reconstruction using GPU acceleration methods. Results showed that the proposed projector provides better or equivalent contrast, lower noise and high resolution compared to the other projectors, with a fast reconstruction time. Our future work will consist of optimizing this new projector for the GPU architecture and validate it on clinical data.

REFERENCES

[1] A. Rahmim, J.-C. Cheng, S. Blinder, M.-L. Camborde, and V. Sossi, "Statistical dynamic image reconstruction in state-of-the-art high-resolution pet," *Physics in medicine and biology*, vol. 50, pp. 4887–4912, 2005.

[2] J. Qi, R. Leahy, S. Cherry, A. Chatziioannou, and T. Farquhar, "High-resolution 3d bayesian image reconstruction using the micropet small-animal scanner," *Physics in medicine and biology*, vol. 43 , Issue 4, pp. 1001 – 1013, 1998.

[3] R. Siddon, "Fast calculation of the exact radiological path length for a three-dimensional ct array," *Medical Physics*, vol. 12, pp. 252–257, 1985.

[4] J. Cui, G. Pratx, S. Prevrhal, B. Zhang, L. Shao, and L. C.S., "Measurement-based spatially-varying point spread function for list-mode pet reconstruction on gpu," in *IEEE Nuclear Science Symposium Conference Record*, 2011.

[5] J. Cui, G. Pratx, S. Prevrhal, and C. S. Levin, "Fully 3d list-mode time-of-flight pet image reconstruction on gpus using cuda," *Medical Physics*, vol. 38, pp. 6775–6786, 2011.

[6] J. Bert and D. Visvikis, "A fast cpu/gpu ray projector for fully 3d list-mode pet reconstruction," in *IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 4126 – 4130.

[7] A. Reader, K. Erlandsson, M. Flower, and R. Ott, "Fast accurate iterative reconstruction for low-statistics positron volume imaging," *Physics in medicine and biology*, vol. 43, pp. 835–846, 1998.

[8] L. Parra and H. Barrett, "List-mode likelihood: Em algorithm and image quality estimation demonstrated on 2-d pet," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 228–235, 1998.

[9] E. Gonzalez, J. Cui, G. Pratx, M. Bieniosek, P. Olcott, and C. Levin, "Point spread function for pet detectors based on the probability density function of the line segment," in *IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 4386 – 4389.

[10] Y. Chen and S. Glick, "Determination of the system matrix used in list-mode em reconstruction of pet," in *IEEE Nuclear Science Symposium Conference Record*, 2007, pp. 3855 – 3858.

[11] F. Lamare, A. Turzo, Y. Bizais, C. C. L. Rest, and D. Visvikis, "Validation of a monte carlo simulation of the philips allegro/gemini pet systems using gate," *Physics in medicine and biology*, vol. 51, pp. 943–962, 2006.

[12] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, T. Frisson, L. Grevillot, L. Guigues, L. Maigne, C. Morel, Y. Perrot, N. Rehfeld, D. Sarrut, D. Schaart, S. Stute, U. Pietrzyk, D. Visvikis, N. Zahra, and I. Buvat, "Gate v6: a major enhancement of the gate simulation platform enabling modelling of ct and radiotherapy," *Physics in medicine and biology*, vol. 56, pp. 881 – 901, 2011.

[13] M. A. Lodge, A. Rahmim, and R. L. Wahl, "Simultaneous meausrement of noise and spatial resolution in pet phantom images," *Physics in medicine and biology*, vol. 55, pp. 1069–1081, 2010.

High Performance Parallel Beam and Perspective Cone–Beam Backprojection for CT Image Reconstruction on Pre–Production Intel® Xeon Phi™

Matthias Baer and Marc Kachelrieß

***Abstract*—With the Xeon Phi™ coprocessor Intel® recently introduced a new many core hardware platform. In this work we want to present some of the basic hardware properties of Xeon Phi™ as well as performance results of optimized implementations of a parallel beam and a perspective cone–beam backprojection. Results are compared to those achieved on the CPU and the GPU.**

Keywords: High performance computing, optimization, vectorization, parallel computing, image reconstruction

I. Introduction

Whenever a new compute architecture is introduced, existing code must be redesigned and optimized to get the highest performance out of the new hardware. Recently Intel® introduced a many core computation platform named Xeon Phi™. The fact that Xeon Phi™ is a freely programmable and not a special purpose processor makes it very attractive to high–end applications such as medical imaging. The aim of this investigation is to discuss some of the hardware features of Xeon Phi™ and to implement an optimized parallel beam and a perspective cone–beam backprojection algorithm for Xeon Phi™. Performance results on Xeon Phi™ will be compared to the performance of CPU– and GPU–optimized versions of the algorithms executed on a state–of–the–art CPU and GPU.

II. Materials and Methods

A. Xeon Phi™

1) Hardware Properties: The Xeon Phi™ performance measurements presented in this paper were conducted on a pre–production engineering hardware sample, stepping B0, software version beta. Xeon Phi™ is a newly developed coprocessor that is connected via the PCIe bus to a host PC, similar as a GPU. However, the hardware layout of Xeon Phi™ is not similar to that of a GPU at all. Xeon Phi™ consists of many Pentium–like x86 cores — our Xeon Phi™ has 61 cores each running at 1.2 Ghz — which are connected via a ring bus for data transfer between the different cores. Each of the cores has its own 32 kB L1 and a 512 kB L2 cache. All cores share their L2 cache via a ring network. Additionally each of the cores features a 512 bit vector unit which is accessed by a new instruction set. This means that 16 32 bit floating point values (floats) or eight 64 bit floating point values (doubles) can be processed with a single instruction on Xeon Phi™.

Besides the high level of parallelism Xeon Phi™ comes with 8 GB of system RAM which is sufficient also for algorithms that have higher demands in terms of memory as for example iterative image reconstruction algorithms.

2) Programming Model and Vector Instruction Set: Since Xeon Phi™'s processor cores are based on the x86 Pentium architecture there is no need for a special programming language, like CUDA on NVIDIA GPUs, to implement algorithms for Xeon Phi™. Code for Xeon Phi™ can be developed using either C/C++ or Fortran. Regions in the program which should be offloaded and executed on Xeon Phi™ are marked by a simple pragma directive. Within the offload region C/C++ respectively Fortran can be used. Parallelism can be added by using OpenMP. Due to this simple programming model existing code can be easily ported to Xeon Phi™. Of course adding only the offload pragma does not ensure that the code is running efficiently and fast on Xeon Phi™. A key factor to achieve a high performance is vectorization. Thereby the instruction set that operates on the 512 bit registers is more flexible and has additional instructions as compared to the SSE and AVX vector instruction sets on standard Intel® CPUs. For the implementation of the parallel beam and perspective cone–beam backprojection especially the new fused multiply–add and the scatter / gather functionality of the Xeon Phi™ instruction set helped to ease or even to enable vectorization. The fused multiply–add functionality for example can be used to save instructions when implementing linear interpolation. The gather instruction on the other hand allows to load data from arbitrary memory locations into one register and the scatter instruction can be used to store register entries to arbitrary memory locations. This scatter / gather functionality enables the vectorization of algorithms that are, due to irregular data patterns, unvectorizable on standard CPUs since their SSE or AVX instruction sets support only load and store from contiguous and well aligned memory locations.

Dr. Matthias Baer, Prof. Dr. Marc Kachelrieß: German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
Corresponding author: matthias.baer@dkfz.de

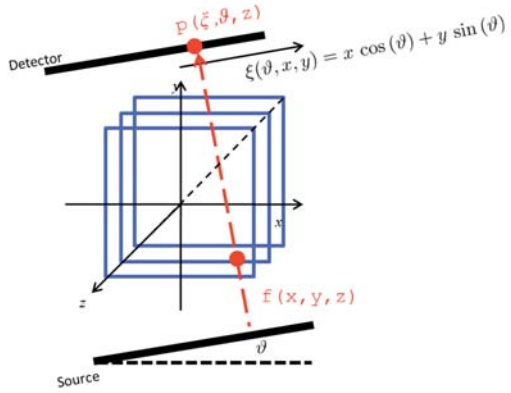


Fig. 1. Geometry of the parallel beam backprojection.

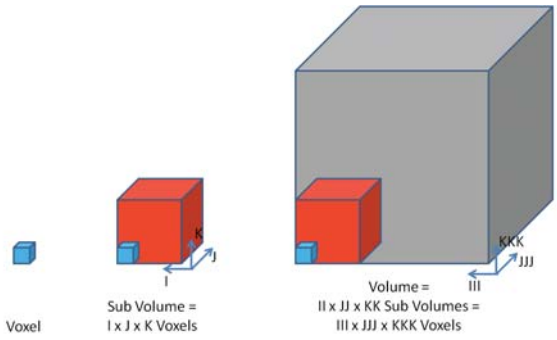


Fig. 2. Volume layout with loop tiling. The volume is tiled in sub volumes in order to increase the data locality and to reduce cache misses

In addition Xeon Phi™ also supports a 16 bit floating point (half) data format. Data that are represented by halfs in memory can be converted to floats while loading data into the register. As long as the reduction in accuracy does not impair the results, using halfs instead of 32 or 64 bit floating point (floats/doubles) values is an option to speed–up bandwidth–limited algorithms [1].

B. Parallel Beam Backprojection

We consider 2D parallel beam backprojections (figure 1) of the type

$$f(x,y) = \int p\big(\vartheta, \xi(\vartheta,x,y)\big)\, d\vartheta$$

Here, ϑ is the projection angle, $p(\vartheta,\xi)$ are preprocessed and convolved projection values, and $f(x,y)$ is the image. The function $\xi(\vartheta,x,y)$ is the distance of the origin to the ray through the point (x,y) for projection angle ϑ. It is given by

$$\xi(\vartheta,x,y) = x\cos\vartheta + y\sin\vartheta \ .$$

During the backprojection a mapping from continuous detector coordinates ξ to discrete detector indices must be done. For the performance evaluation on Xeon Phi™ we investigated two interpolation methods: Nearest neighbor (NN) and linear interpolation (LI). The generalization of the parallel beam backprojection from 2D to 3D is straight forward:

$$f(x,y,z) = \int p\big(\vartheta, \xi(\vartheta,x,y), z\big)\, d\vartheta.$$

Since the function $\xi(\vartheta,x,y)$ does not depend on the z–coordinate of the volume, reconstructing a 3D volume in parallel beam geometry is equal to the reconstruction of several independent image slices (figure 1). The fact that there is not data dependency between different slices offers the possibility to vectorize the parallel beam backprojection, i.e. to backproject several slices simultaneously. If the loop over the z–coordinate is chosen as the innermost loop 16 image slices can be backprojected simultaneously by using the 512 bit vector instruction set of Xeon Phi™.

Besides vectorization we applied a second optimization technique. To enhance data locality the projection data and the volume are divided into sub volumes and sub projections (figure 2). Assume that the volume consists of III×JJJ×KKK voxels and is divided into II×JJ×KK sub volumes each having I×J×K voxels. The projections are divided into NN sub projections each holding N projections whereby the total number of projections is given by NNN=NN×N. The internal loop order of the parallel beam backprojection with loop tiling is now as follows: First a given sub volume is reconstructed from the first N projections. Then the next sub volume is reconstructed from these N projections. This procedure is repeated until the first N projections are backprojected into the full volume. Afterwards the process is repeated for the next N projections. The backprojection is finished if all NN sub projections are backprojected into the full volume. Dividing the volume and the projection data into smaller sub volumes and sub projections reduces the number of cache misses since due to the enhanced data locality, data is more likely already present in the cache when it is actually needed for computation. The reduction of cache misses is essentially to achieve a high performance in case of the parallel beam backprojection since this algorithm is typically bandwidth–limited. To reduce the pressure on system bandwidth we also tested the impact of using a data format with reduced memory needs. As described above Xeon Phi™ supports a 16 bit floating point format (half). Using halfs instead of floats reduces the pressure on the memory bandwidth by 50% which may result in a doubled performance when switching from floats to halfs iff the algorithm is purely bandwidth–limited. In reference [1] it was shown that using halfs instead of floats does not decrease image quality of reconstructed images.

C. Perspective Cone–Beam Backprojection

A possible way to reconstruct 3D volumes $f(x,y,z)$ from projection data in perspective cone–beam geometry is the Feldkamp algorithm [2]. In this paper we want to focus on the backprojection step of a cone–beam image reconstruction and therefore we assume that projection data

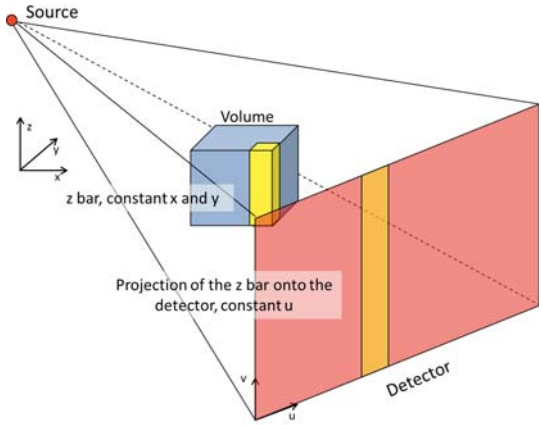


Fig. 3. Geometry of the perspective cone–beam backprojection. The projection of a z bar results in constant u coordinates on the detector.

p are already preprocessed and convolved with a convolution kernel. Making this assumption the backprojection equation reads

$$f(x,y,z) = \int w^2 \, p\big(\alpha, u, v\big) \, d\alpha.$$

Here α is the projection angle, u and v are detector coordinates and w^2 is the distance weight. The detector coordinates and the distance weight can be expressed in terms of the coefficients c_{ij} that define the perspective transform from the volume on the detector.

$$\begin{aligned} u(\alpha,x,y,z) &= (c_{00}x + c_{01}y + c_{02}z + c_{03})w(\alpha,x,y,z) \\ v(\alpha,x,y,z) &= (c_{10}x + c_{11}y + c_{12}z + c_{13})w(\alpha,x,y,z) \\ w(\alpha,x,y,z) &= (c_{20}x + c_{21}y + c_{22}z + c_{23})^{-1} \end{aligned}$$

We here assume that the v–axis of the detector is aligned parallel to the volume's z–axis (figure 3). This yields $c_{02} = c_{22} = 0$ [3] and now the projection coordinate u is the same for all voxels in z–direction for given coordinates x and y (figure 3). Additionally the weights w^2 are also the same for all voxels in z–direction with constant x and y. This fact can be used to avoid divisions in the innermost loop of the backprojection. If choosing this loop to run over the z–coordinate of the volume the weights w^2 can be precalculated and since u remains the same for constant x and y, the calculation of the projection coordinate v in the innermost loop results in

$$v(\alpha,x,y,z) = o + sz.$$

The slope s and the offset o can be expressed by

$$\begin{aligned} s &= w(\alpha,x,y) \cdot c_{12} \\ o &= w(\alpha,x,y) \cdot (c_{10}x + c_{11}y + c_{13}) \end{aligned}$$

and remain constant for voxels with constant x and y.

As in the case of the parallel beam backprojection we used vectorization and loop tiling optimization (figure 2). Also for the perspective cone–beam backprojection the loop over the z–coordinate was chosen as innermost loop. Since the slope s and the offset o do not depend on z, detector indices v for several succeeding voxels of one z bar can be computed simultaneously. Using Xeon Phi™'s 512 bit vector unit, 16 detector indices v can be computed simultaneously. Due to the relationship between z–coordinates and detector coordinates v the data access to the projection data p for voxels of a given z–bar is irregular. This fact hinders the vectorization of the perspective cone–beam backprojection on CPUs since their vector instruction sets, may it be SSE or AVX, support only loads and stores from contiguously and well aligned memory locations. Xeon Phi™ in contrast, has gather / scatter support so even data from arbitrary spread memory locations can be loaded into a single vector register. Using this functionality now allows for vectorization by processing the projection data according to 16 subsequent voxels of a given z–bar simultaneously.

The loop tiling for the perspective cone–beam backprojection was realized in the same way as in the case of the parallel beam backprojection and also the usage of halfs instead of floats to represent the volume and the projection data was investigated as well as the two different interpolation methods NN and LI.

D. CPU and GPU References

The reference CPU performance measurements were done on the host PC of Xeon Phi™. In our case the host PC is equipped with 2 Intel® Xeon E5–2670 processors. The clock speed of the processors is 2.6 GHz and each processor has 6 cores. This makes a total of 24 threads running in parallel since hyperthreading was enabled. The CPU implementation of the parallel beam and perspective cone–beam backprojection were optimized to the same level as the Xeon Phi™ implementations [4]. The vectorization of the parallel beam backprojection was done using the AVX instruction set. The perspective cone–beam backprojection is unvectorizable on the CPU due to the missing gather / scatter support.

The reference GPU performance measurements were done on a NVIDIA Quadro 6000 GPU. Also in the GPU case the implementation of the parallel beam and perspective cone–beam backprojections were optimized to a high level, similar to the optimization level of the Xeon Phi™ implementation and the CPU implementation [5, 6].

E. Test Cases and Performance Metric

The test cases for the performance measurements were chosen so that they meet typical problem sizes for CT image reconstruction. For the parallel beam backprojection we chose a volume having 512^3 voxels, a detector size of 512^2 and backprojected 512 projections. For the perspective cone–beam backprojection we chose a volume size of 512^3, a detector having 1024^2 pixels and backprojected 720 projections.

As described above the volume and the projections were divided into smaller sub volumes and sub projections (figure 2). The optimal size of these sub volumes and sub projections was determined empirically by choosing the setting with the best performance.

	Xeon Phi™ (floats)	Xeon Phi™ (halfs)	CPU (floats)	GPU (floats)
NN	91 GUPS	165 GUPS	106 GUPS	25 GUPS
LI	58 GUPS	81 GUPS	51 GUPS	25 GUPS

TABLE I
PERFORMANCE OF THE PARALLEL BEAM BACKPROJECTION ON XEON PHI™, CPU AND GPU FOR NEAREST NEIGHBOR (NN) AND LINEAR INTERPOLATION (LI).

	Xeon Phi™ (floats)	Xeon Phi™ (halfs)	CPU (floats)	GPU (floats)
NN	53 GUPS	59 GUPS	17 GUPS	25 GUPS
LI	27 GUPS	31 GUPS	7 GUPS	25 GUPS

TABLE II
PERFORMANCE OF THE PERSPECTIVE CONE–BEAM BACKPROJECTION ON XEON PHI™, CPU AND GPU FOR NEAREST NEIGHBOR (NN) AND LINEAR INTERPOLATION (LI).

The performance metric we used to quantify our results is Giga Updates Per Second (GUPS). Thereby a single update consists of loading a projection value, reading a value from the volume, adding — with or without interpolation — both values and storing the result back to the volume. So in case of the parallel beam backprojection $512^3 \times 512 = 64$ Giga Updates (GU) are needed to backproject all projection data into the volume. For the test setup of the perspective case the total number of Giga Updates is 85.

III. RESULTS

In table I the maximum performance values for the parallel beam backprojection on Xeon Phi™ are depicted. For comparison table I also shows the results for the CPU and the GPU. The performance of Xeon Phi™ is at least by a factor of two higher as compared to the GPU for all tested cases. Here it must be noted that the GPU values given in table II are those of the perspective cone–beam backprojection. The NVIDIA Quadro 6000 GPU was not available to us the time we wrote this paper but previous investigations [5] showed that the performance of parallel beam and perspective cone–beam backprojection is the same on the GPU. As compared to the CPU Xeon Phi™ shows about the same performance when using floats for the volume and the projection data. A big impact on performance can be observed when switching from floats to halfs. Doing so yields a speed–up of about 80% for the NN case and 40% for the LI case. The lower impact of the reduction in data size on performance in the LI case can be explained by the fact that here the ratio of memory operations to computations drops as compared to NN and therefore the reduction of bandwidth needs has a lower impact on the final overall performance.

The results for the perspective cone–beam backprojection are given in table II. As compared to the CPU the performance of Xeon Phi™ is at least by a factor of 3 higher. As compared to the GPU, Xeon Phi™ is about a factor of 2 faster in the NN case and reaches about the same performance in the LI case. The impact of the data format on the performance is not that high as in the parallel case. When switching from float to half the performance increases is only by about 10% for both interpolation methods. The reason for this may be the fact that the perspective backprojection has a lower ratio of memory operations to computations and therefore the impact of reduced bandwidth needs is not as strong as for the parallel case.

IV. CONCLUSION

We here implemented a parallel beam and a perspective cone–beam backprojection algorithm on the Xeon Phi™ coprocessor and compared the performance values to the results achieved with a state of the art CPU and GPU.

When looking at the results it shows up that both the CPU and the GPU have their favorites in terms of backprojection algorithms. The CPU is faster than the GPU in the parallel case while the GPU is faster in the perspective case. Xeon Phi™ on the other hand is at least competitive with both architectures even for their best cases and outperforms them, at least slightly, if using the half format for the volume and the projection data. Comparing the performance results of Xeon Phi™ with those of the GPU it shows up that Xeon Phi™ is faster than the GPU in all cases except when linear interpolation is used. The reason for this is that linear interpolation is implemented in hardware on the GPU and therefore comes for free without extra computations as compared to nearest neighbor. Nevertheless although having no hardware support for linear interpolation Xeon Phi™ reaches the same performance as the GPU, or slightly outperforms the GPU when using halfs, even when linear interpolation is used. In conclusion we think that due to its good performance and due to the fact that porting existing C/C++ code to Xeon Phi™ can be achieved with only minor modifications, Xeon Phi™ may be a good workhorse in the fields of medical imaging.

V. ACKNOWLEDGMENTS

This study was supported by a grant from the Intel® Corporation. Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

REFERENCES

[1] C. Maaß, M. Baer, and M. Kachelrieß. CT image reconstruction with half precision floating–point values. *Med. Phys.*, 38(2):691–700, 2011.
[2] L.A. Feldkamp, L.C. Davis, and J.W. Kress. Practical cone–beam algorithm. *Journal of the Optical Society of America*, 1(6):612–619, June 1984.
[3] M. Kachelrieß, M. Knaup, and O. Bockenbach. Hyperfast parallel–beam and cone–beam backprojection using the Cell general purpose hardware. *Med. Phys.*, 34(4):1474–1486, April 2007.
[4] M. Knaup and M. Kachelrieß. Acceleration techniques for 2D parallel and 3D perspective forward– and backprojections. *Proceedings of the 9th Fully3D*, pages 45–48, 2007.
[5] M. Knaup and M. Kachelrieß. GPU–based parallel–beam and cone–beam forward– and backprojection using CUDA. *IEEE Nuclear Science Symposium Conference Record*, pages M10–354, 2008.
[6] S. Sawall, L. Ritschl, M. Knaup, and M. Kachelrieß. Performance comparison of OpenCL and CUDA by benchmarking an optimized perspective backprojection. *Proceedings of the 11th Fully3D and the 3rd HPIR Workshop*, 2011.

Cloud X: A Platform as a Service for CT Reconstruction Research and Development

Eric Papenhausen, Ziyi Zheng, and Klaus Mueller

***Abstract* – Many CT reconstruction algorithms, especially in iterative CT, are constructed from a few common algorithms (e.g. backprojection and forward projection). These algorithms often dominate the computation cost during CT reconstruction. GPUs can provide an order of magnitude speed-up over conventional CPU implementations. However, without the proper hardware or willingness to develop the specialized software, these speedups will remain unexploited. In addition to implementing these common algorithms, sharing research can also be a burden. Assuming one is able to get the software to run someone else's experiments, one can spend an entire work day tracking down the appropriate dependencies and modifying hard coded file paths to run it on a different machine. These obstacles hinder productivity and slow research. In this paper, we present a cloud computing framework that aims to make research in the medical imaging domain more efficient by providing a number of common GPU accelerated algorithms and allows for efficient sharing of research through a virtual workspace. It also has an interface that allows users to present and analyze their research.**

***Index Terms*—Cloud Computing, GPU, CT reconstruction**

I. INTRODUCTION

With the introduction and rapid adoption of GPUs to the medical imaging domain [6][8][9][13], it is becoming increasingly important to leverage the processing power of GPUs to make the leap from research to clinical use. Developing software on GPUs, however, can seem like a daunting task. It not only requires the appropriate hardware and software tools, but a shift in the mindset of how a program operates. This can be especially difficult for theoretical researchers, who have to rely on someone else to accelerate their algorithms on graphics hardware.

Furthermore, comparing different CT reconstruction methods can require a re-implementation of a previous technique. A popular iterative CT reconstruction algorithm is ASD-POCS [10]. When a new iterative CT algorithm is proposed, it is often compared to ASD-POCS and other techniques [1][2][3] to demonstrate the validity of the new algorithm. Researchers will often have to re-implement multiple reconstruction algorithms to compare to their new method; then mention that they are unsure of their implementation of the old methods.

Most research today is relatively one sided (i.e. the authors explain how an experiment was set up and then present results), but there is no easy way for the reader to validate and further explore a presented algorithm. There are a number of parameters that can affect the quality of a CT reconstruction algorithm (i.e. number of projections, dose per projection, anatomy and pathology, etc.), but only a subset of these parameters are typically presented.

In this paper, we present a cloud computing framework that will be able to solve these problems by providing three main services. First, it will provide a virtual workspace which will facilitate the sharing of research and contain a number of pre-loaded GPU accelerated CT reconstruction algorithms. Second, it will allow remote execution of GPU accelerated algorithms, thus allowing users to take advantage of graphics cards without having to purchase the hardware or develop GPU accelerated applications. Finally, it will provide an interface that will allow authors to post their research and allow users to replicate experiments and explore the parameter space through their web browser.

We begin in section II by presenting an interface that we have developed to allow users to explore the parameter space of an iterative CT reconstruction algorithm through their web browser. The rest of the paper will then be focused on how to build a scalable framework around this interface and the other services that this framework will provide. In section III, we will discuss the current technology that will allow us to build this cloud computing framework. Section IV will describe the architecture of our framework and introduce the concept of the virtual workspace. In section V, we revisit our web interface and describe how it will operate in the cloud. Section VI presents a library that will allow remote execution of GPU accelerated algorithms. Finally, section VII concludes the paper.

II. WEB INTERFACE

We have developed a prototype of a web interface that allows users to explore the parameter space of an iterative CT reconstruction algorithm. More specifically the algorithm is OS-SIRT [11] using a bilateral filter (BLF)[12] for regularization. The bilateral filter is defined in the following equations, where x is the location of smoothing, and t is an offset. The BLF is essentially a Gaussian that falls off as a function of both spatial and value deviation.

Eric Papenhausen, Ziyi Zheng (now with Amazon.com, Seattle), and Klaus Mueller are (or have been) with the Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University, Stony Brook, NY 11777 USA (phone: 631-632-1524; e-mail: {epapenhausen, zizhen, mueller}@cs.sunysb.edu).

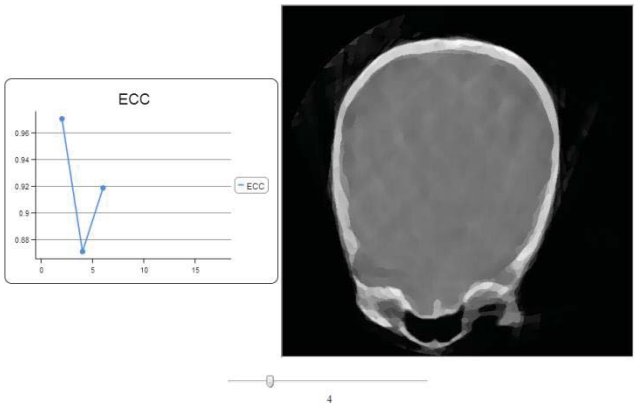


Figure 1. A view of the web interface that we have developed.

$$\boldsymbol{\alpha_{BLF}(x,t,f) = c_d(t)s_r(f(x),f(x+t))}$$
$$\boldsymbol{c_d(t) = G_{\sigma_d}(\|t\|)} \quad (1)$$
$$\boldsymbol{s_r\big(f(x),f(x+t)\big) = G_{\sigma_r}(|f(x)-f(x+t)|)}$$

Where $G_\sigma(\boldsymbol{x})$ is the Gaussian kernel:

$$\boldsymbol{G_{\sigma_r}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{x^2}{2\sigma^2})} \quad (2)$$

We allow the user to explore the filter's parameter space interactively using a slider to determine the amount of smoothing. The slider changes the parameters of the BLF to control the smoothing such that the value chosen by the slider determines σ_r and σ_d (i.e. $\sigma_r = \sigma_d$).

We implemented this interface using HTML and JavaScript on the front end, PHP as the server side scripting language, and finally using C and CUDA to handle the CT reconstruction. The act of moving the slider triggers a request to the web server. The web server then writes the parameter value, chosen by the slider, to a file. A simple C program waits for this file to be updated, and when it is, it calls the OS-SIRT function with the appropriate parameters, and waits for the volume to be reconstructed. The central slice of the reconstructed volume along with its E-CC score is then passed back to the webserver to be displayed to the user. Fig. 1 shows the web interface that the user sees. The graph on the left plots the quality metric curve evolved with each slider update and the right shows the current image.

Currently, this setup is running on a single desktop machine with a NVIDIA GeForce GTX 480. It takes around 11 seconds for each reconstruction (i.e. 512^3 volume for OS-SIRT 10 with 60 projections and a BLF for regularization). We have available to us, however, a GPU cluster that contains 8 NVIDIA Tesla M2050 GPUs. The framework that we present in the rest of this paper is designed to run on this GPU cluster. We expect the reconstruction time to drop to between 1.3 and 1.8 seconds. This, however, does not mean that the user will have to wait 1.3-1.8 seconds every time he moves the slider, since we will include a number of optimization strategies, presented throughout this paper.

III. CLOUD COMPUTING

With the popularization of cloud computing, a number of open source technologies have been developed to support the infrastructure and system level requirements that are required by cloud computing frameworks [4][7]. Openstack in particular is a popular cloud operating system. It contains a number of networking and storage solutions that are required by the typical cloud computing infrastructure. It also allows for the creation and management of multiple virtual machines (i.e. software that emulates the computer architecture of a real machine). This is particularly important for efficient sharing of computational resources. Unfortunately, openstack currently does not contain any resources for efficient GPU virtualization (i.e. sharing of GPU resources).

Gvirtus is an open source GPU virtualization technology. It allows multiple virtual machines to share the computational resources provided by one or more GPUs. To a user operating on a virtual machine, however, it appears as if there is one dedicated GPU. This abstraction allows for efficient GPU sharing, while still providing the user an interface he is familiar with.

Another important tool for cloud computing is the network file system[5] (NFS). This allows many users to connect to a single file system concurrently. To the user, however, the NFS looks like a regular file system. This allows for efficient sharing of information.

IV. FRAMEWORK ARCHITECTURE

The cloud computing framework we are currently developing and have initial results for utilizes all of the technology presented in section III (i.e. openstack, gvirtus, and NFS) as well as custom improvements to make the overall system more efficient. The virtual workspace is designed as a "platform as a service" (PaaS) and will allow users to work in the cloud through their web browser. Each user will have a separate virtual machine assigned to them and will contain a number of relevant GPU accelerated CT reconstruction algorithms pre-configured on their workspace. There will be one virtual machine assigned to handle the remote execution of GPU accelerated CT reconstruction algorithms. Finally, there will be one virtual machine to function as a web server, and will be configured to allow users to upload their implementations, allowing it to be executed through a web browser. Since every service is being handled through virtual machines, the 8 GPUs in our cluster can be efficiently shared using gvirtus.

Gvirtus consists of two segments, a front end, located on the virtual machine, and a back end, located on the host operating system. The front end provides a library that acts as a wrapper to many CUDA functions. The front end redirects calls to these functions to the back end. The back end executes the CUDA operations and returns the results back to the front end on the virtual machine.

The network file system will contain a number of commonly used dependencies (i.e. CUDA libraries, include files, etc.) as well as a number of datasets available for CT reconstruction. It will also include a public folder for each user. The public folder allows a user to share his research with the rest of the community. Other users can easily download software from a public folder into their workspace. Since the dependencies are in a shared file system, the software will run without requiring additional modifications.

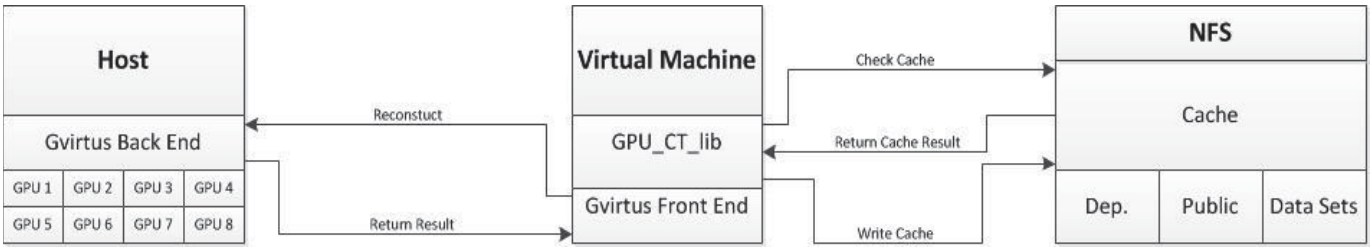


Figure 2. An illustration of the framework presented in this paper. When a function from the "GPU_CT_lib" is called, the system first checks the NFS to see if the result lies in cache. If so, the result is returned. On a cache miss, the appropriate reconstruction function is redirected to the back end (i.e. host machine) through Gvirtus; the result is reconstructed and returned. That result is then written to the cache on the NFS.

The network file system will also contain a block of space to act as cache. A feature vector, containing all the parameters of a previously reconstructed volume (i.e. dataset, number of projections, etc.) and the volume itself, will be recorded. With this strategy, we can alleviate GPU resource contention by only executing CT reconstruction algorithms for which we do not already have the results. Figure 2 shows how this strategy interacts with the rest of the system. We will, however, provide a "no-cache" option for situations where obtaining accurate timings of an algorithm are important.

V. WEB INTERFACE REVISITED

A critical aspect of performing research is presenting the results. This is usually fulfilled by presenting an image of a slice of a reconstructed volume followed by a discussion of how it compares to a gold standard. The presented slice, however, is only one of many possible reconstructions. Results can vary greatly depending on the parameters that are chosen for the reconstruction. Moreover, the quality of the algorithm can differ depending on the anatomy of the dataset that is being reconstructed. By providing an interface that allows readers to access the research in question, experiments can be easily replicated and the algorithm can be thoroughly evaluated.

In section II, we described the details of a prototype interface that addressed these problems. This, however, was specific to OS-SIRT and the BLF. To make this interface more useful, there needs to be a mechanism that allows researchers to attach their project to this interface.

In order for researchers to utilize this interface, however, the process of uploading algorithms needs to be simple. The researcher will need to implement certain methods that will be called when a parameter is changed via a slider. These methods will provide the new parameter settings as an argument. The researcher will simply set the parameters of his algorithm based on the values provided by the web interface, and execute his algorithm. Figure 3 shows a pseudo-code example of the functions that will be needed by the web interface.

```
void sliderSetParam1(float val) { sigma = val; }

void sliderSetParam2(float val) { lambda = val; }

void execute() { . . .}
```

Figure 3. A pseudo-code example demonstrating the functions that will need to be defined for the web interface. The functions with the "slider" prefix will simply set the parameters of the algorithm. The execute function will actually perform the algorithm

Once the appropriate functions are implemented, their program will need to be compiled as a dynamic library. The compiled code will be placed in a directory that is accessible by the web server. By compiling the project as a dynamic library, the web server can call the functions it needs to, without having to restart the server itself. This is important because multiple users will be sharing the same web server. When the dynamic library is placed in the appropriate directory, users will be able to visit a web page specific to the project in question, and will be able to explore and execute the algorithm in real time.

There are some performance issues since there will still be a noticeable delay from when the slider is changed, to when the reconstruction is complete. In section II, we anticipated that the reconstruction time would be between 1.3-1.8 seconds using our cluster of 8 GPUs. With multiple users, this can take much longer since a user will have to wait for a GPU to be free before reconstruction can take place. Some extra optimization strategies are required to reduce GPU resource contention and ensure that web interface remains interactive.

One strategy that we have already presented is the use of cache. Instead of immediately reconstructing the volume, we check to make sure that it has not already been reconstructed. If it has, we simply return the results without performing any extra computation. This is particularly useful with the slider, as users will often move the slider back and forth, reviewing parameter settings they have already seen.

Another strategy is to only reconstruct a subset of the volume. Since only the central slice is displayed to the user, most of the three-dimensional volume is never seen. We estimate that we can reconstruct a 512x512x8 subset of the volume in 0.3-0.7 seconds using our 8 GPU cluster. This strategy combined with caching will greatly reduce the strain on the GPU server while still providing the user with a real-time interface in which he can explore the parameter space of a CT reconstruction algorithm.

VI. REMOTE EXECUTION

The virtual workspace provides a number of resources that are beneficial when starting a new project. This option may not be ideal, however, when maintaining an existing project. A researcher may still take advantage of the processing power of GPUs by using our remote execution library. This library aims to provide a simple method for executing a number of commonly required GPU accelerated algorithms (i.e. backprojection, forward projection, etc.) in the cloud.

The remote execution library will provide a number of wrapper methods to the GPU accelerated algorithms that are provided in the virtual workspaces. These methods operate by acting as a liaison between the user's machine, and the virtual machine in the cloud responsible for processing remote execution requests.

```
cloudTCPConnect();
cloudAllocMem(vol, size);
cloudAllocMem(volp1, size);
cloudAllocMem(proj, projection_size);
for each iteration
    cloudProject_GPU(vol, proj);
    cloudCorrect_GPU(proj, "dataset");
    cloudBackproject_GPU(volp1, proj);
    cloudBLF_GPU(volp1);
    cloudVoxelUpdate_GPU(vol, volp1, lambda);
end for
cloudTCPDestroy();
```

Figure 4. An iterative CT implementation using the remote execution library. Functions with the prefix cloud will be executed on the cloud.

Figure 4 shows a pseudo-code example of how users can interact with the remote execution library. On the server side, a virtual machine will be tasked with listening and waiting for a connection request from a client. On the client side, a TCP connection is initiated with the server. This allows for effective communication between the client and the server. A call to execute a GPU accelerated algorithm by the client sends a signal to the server, indicating the parameters the user requested (i.e. algorithm, dataset, etc.). The server then executes the requested function in a similar manner illustrated in Figure 2. The client can then request the results to be transferred from the cloud to the user's machine.

We recognize that most CT reconstruction pipelines (e.g. iterative CT), however, can be highly integrated (i.e. the results of the "project" section are given as input in the "backproject" section). To reduce the number of memory copies between the client and server, we provide a mechanism to allow the user to allocate space on the cloud. With this functionality, the user can allocate an array to store the results of the "project" function, and pass it as input to the "backproject" function. This reduces costly data transfers between the client and server.

If, however, one component of this pipeline is missing in our library then it becomes useless; since any performance gained by exploiting GPUs will be wiped out by the cost of transferring data to and from the server on every iteration. To avoid this situation, we let users extend the remote execution library by allowing them to upload custom implementations to the remote execution server. A simple addition to the wrapper library, is then required. In this fashion, users can easily expand the library to suit their needs.

With the remote execution library, users can exploit the massive speed-ups GPUs provide without the burden of developing highly specialized software. This library will be primarily useful for those looking to quickly replace sequential CPU implemented CT algorithms with GPU accelerated implementations.

VII. CONCLUSIONS

In this paper, we presented an interface for dynamically exploring the parameter space of an iterative CT reconstruction algorithm through a web browser. We then explained how we can build a scalable framework to allow many users to interact with this interface in a cloud environment while handling practical issues related to resource contention. The technology developed for cloud based computing infrastructures are quite powerful and allow us to do so much more than simply host web interface for exploring iterative CT reconstruction parameters.

This cloud framework can be used to provide resources that are not necessarily available (i.e. hardware and software) to many researchers. We address many practical issues in the medical imaging community relating to the sharing and presentation of research. The cloud computing framework presented in this paper will be a powerful force for good in the medical imaging community.

REFERENCES

[1] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," Ultrason. Imaging, 6:81–94, 1984

[2] A. Chambolle, "An Algorithm for Total Variation Minimizations and Applications," J. Mathematical. Imaging and Vision 20(1-2) 89-97 (2004).

[3] P. Gilbert, "Iterative methods for the 3D reconstruction of an object from projections," Journal of Theoretical Biology 76, 105–117 (1972).

[4] G. Giunta, R. Montella, G. Agrillo, and G. Coviello. "A GPGPU transparent virtualization component for high performance computing clouds". In Euro-Par 2010 - Parallel Processing, volume 6271 of Lecture Notes in Computer Science, chapter 37, pages 379–391–391. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2010.

[5] A. Muthitacharoen, B. Chen, and D. Mazieres. A Low-bandwidth Network File System. In Proc. 18th SOSP, Oct. 2001.

[6] Y. Okitsu, F. Ino and K. Hagihara. "High-Performance Cone Beam Reconstruction Using CUDA Compatible GPUs," Parallel Computing,36(2-3):129-141, 2010.

[7] OpenStack LLC, "OpenStack: The Open Source Cloud Operating System," 21-Jul-2012. [Online]. Available: http://www.openstack.org/software/.

[8] E. Papenhausen, Z. Zheng, and K. Mueller, "GPU-Accelerated Back-Projecting Revisited: Squeezing Performance by Careful Tuning," Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Potsdam, Germany, 2011)

[9] H. Scherl, B. Keck, M. Kowarschik, J. Hornegger, "Fast GPU-based CT reconstruction using the Common Unified Device Architecture(CUDA)," IEEE Medical Imaging Conference, 6: 4464-4466, Honolulu, HI, 2007.

[10] E.Y. Sidky and X. Pan, "Image Reconstruction in Circular Cone-Bean Computed Tomography by Constrained, Total Variation Minimization", Phys. Med. Biol. 53 (2008) 4777-4907

[11] F. Xu, W. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard and K. Mueller, "On the Efficiency of Iterative Ordered Subset Reconstruction Algorithms for Acceleration on GPUs," Computer Methods and Programs in Biomedicine 98(3) 261-270 (2010).

[12] W. Xu, K. Mueller, "Evaluating Popular Non-Linear Image Processing Filters for their Use in Regularized Iterative CT," IEEE Medical Imaging Conference, Knoxville, TN, October, 2010

[13] Z. Zheng, K. Mueller "Cache-Aware GPU Memory Scheduling Scheme for CT Back-Projection," IEEE Medical Imaging Conference, Oct. 2010.

Iterative Helical CT Reconstruction in the Cloud for Ten Dollars in Five Minutes

Jeffrey M. Rosen, Junjie Wu, Jeffrey A. Fessler, Thomas F. Wenisch
Department of EECS, University of Michigan

***Abstract*—Iterative statistical X-ray CT reconstruction algorithms can improve image quality for low dose scans. Unfortunately, their clinical utility has been hampered by their enormous computational requirements; typical low-dose reconstructions require about an hour on commercial systems. Most existing parallel implementations use a shared memory programming model, limiting available parallelism. We investigate using a large compute cluster for a penalized weighted least-squares algorithm using ordered subsets (PWLS-OS), scaled to hundreds of cores to accelerate a single helical CT reconstruction problem. Using Amazon's Elastic Cloud Compute (EC2) service, our experimental results show that a typical helical chest scan can be reconstructed in under five minutes at a cost under $10.**

I. Introduction

Model-based iterative reconstruction for X-ray CT can improve image quality and promises to enable X-ray dose reductions compared to conventional filtered back-projection [1]. Such methods use statistical models and imaging system models, improving image quality. The primary drawback of statistical reconstruction methods is their massive computational requirement. Current commercial model-based reconstruction methods can require about an hour to reconstruct a typical helical chest scan. Improving reconstruction times is essential to enable ubiquitous use of low-dose CT.

Researchers are developing reconstruction algorithms that reduce computational requirements and/or converge more quickly. For example, one recent ordered-subsets algorithm reaches its limit cycle in about 20 iterations [2]. Nevertheless, compute-times-per-iteration remain high (several minutes for helical chest CT scans), so matching scanner and reconstruction throughput requires further improvements.

Parallelization can reduce time-per-iteration, leveraging the multiple cores present in modern processors by partitioning computation into multiple simultaneous sub-problems [3]. Most existing parallel implementations share image and sinogram data in a global main memory accessible to all processor cores. Shared memory simplifies parallelization—each core computes a subset of the image/sinogram, and can read from any part of the image/sinogram space with only infrequent synchronization at coarse steps of each iteration. However, shared memory approaches are limited by the number of cores that can be provisioned in a single system—at most a few tens of cores in conventional commodity systems.

In this work, we investigate the alternative of leveraging the scalability of massive compute clusters to apply distributed computing power to image reconstruction. We demonstrate distributed image reconstruction using leased resources from a commercial cloud computing provider. Cloud services provide low-cost, on-demand, commodity computing resources. They are relatively cheap when compared to purchasing a cluster, and can either be used on-demand for flexibility or reserved for exclusive use and further discounted costs. They also provide a low-overhead mechanism for expanding computational power. To increase the number of nodes working on a problem, one need only purchase more cloud compute time (only seconds of setup time).

Although we use the cloud to demonstrate the performance potential of distributed reconstruction, our methods are applicable more generally to all distributed systems. For example, researchers have accelerated reconstruction using the parallelism in graphics processing unit (GPU) accelerators [4]. However, ganging multiple GPUs for greater parallelism presents a significant challenge because they do not share a single global address space, hence data dependencies are problematic. Our methods could be applied to such a system to sub-divide processing across distributed GPUs. Similarly, emerging devices such as the Xeon Phi coprocessor can be ganged together to achieve greater performance and scalability using our approach.

To use distributed computing resources, one must parallelize reconstruction algorithms across compute nodes that do not share a single global memory. We follow the paradigm of many large-scale scientific applications by using explicit *message passing* to exchange updates to sinogram and image data between compute nodes at appropriate synchronization points in the reconstruction algorithm. As we will show, we can easily scale the number of nodes collaborating on a single reconstruction problem until performance is limited by available communication bandwidth; further speedups will require either faster (and more expensive) interconnection networks or innovations to reduce data communication.

Prior efforts to parallelize filtered back-projection over a cluster have used the MapReduce programming model [5, 6], wherein the computation is translated into simple "map" and "reduce" tasks and a runtime system orchestrates communication among these tasks. Though they ease programming, publicly available MapReduce frameworks, such as Hadoop, store intermediate results on disk when communicating, which is extremely inefficient for iterative reconstruction algorithms. To our knowledge, there have been no prior reports of iterative reconstruction of clinical helical CT scans (with thousands of projection views) using hundreds of cores on a commodity cloud computing service. The closest related work is the

JAF supported in part by NIH grant R01 HL 098686 and by equipment donations from Intel.

The authors thank Donghwan Kim for assistance with the algorithms in [2, 8].

investigation by Gregor of an unregularized SIRT algorithm on four 8-core nodes for axial micro CT with 360 views [7].

This paper investigates a penalized weighted least-squares with ordered subsets (PWLS-OS) reconstruction algorithm [8] that distributes computation across several multi-core nodes that communicate via explicit message passing. This approach scales beyond the limits of a single node, so the number of cores working in parallel is limited only by available hardware, communication bandwidth, and cost. We use Amazon's Elastic Compute Cloud (EC2) service to demonstrate the potential of cloud computing environments and show that a 20-iteration reconstruction of a 320-slice helical chest CT scan using 50 nodes (800 cores) requires less than five minutes at a total computing cost under $10. A "private" cloud computing environment (e.g., operated under contract for a large hospital network) might approach similar costs.

II. Methods

A. Background

We compute a reconstructed image $\hat{\mathbf{x}}$ by minimizing a PWLS cost function [1]:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2_{\mathbf{W}} + R(\mathbf{x}), \tag{1}$$

where $\mathbf{y}$ denotes the observed X-ray CT sinogram data, $\mathbf{W}$ denotes a diagonal statistical weighting matrix, $\mathbf{A}$ is the system matrix [9], and $R(\mathbf{x})$ is an edge-preserving regularizer that balances noise and image resolution. We use an $R(\mathbf{x})$ with first-order finite differences between a voxel and its closest 26 neighbors and a Fair edge-preserving potential [10], but our methods can be extended to other regularizers.

The PWLS-OS iterative algorithm [3] involves four steps: forward projection, back projection, regularization, and image update. Figure 1 depicts the algorithm graphically. The first step involves forward projecting the estimate $\mathbf{x}_n$ at the nth iteration and calculating the weighted sinogram residual:

$$\mathbf{r}_n = \mathbf{W}(\mathbf{A}\mathbf{x}_n - \mathbf{y}). \tag{2}$$

The calculations for each residual are independent, so they can be arbitrarily reordered or parallelized [9]. Iterating along the scan axis in the innermost loop enables reuse of beam geometry calculations. Only the projection views within a given subset are projected in a given sub-iteration, so methods that use fewer subsets, and hence more views per subset, *e.g.*, [2], provide more opportunity for parallelism.

Back projection applies the transpose of the system matrix

$$\mathbf{b}_n = \mathbf{A}'\mathbf{r}_n. \tag{3}$$

Back projection must occur after forward projection because of its dependency on $\mathbf{r}_n$. In principle, one can backproject the residual into every voxel independently, allowing massive parallelism over voxel space. In practice it is again more efficient to have an inner loop along the axial direction for each thread [9]. This strategy leads to parallelization across the $\approx 512^2$ voxels in a single transaxial slice, which still allows for tens of thousands of threads.

The regularization step calculates the gradient ∇ of the regularizer at the current image $\mathbf{x}_n$:

$$\mathbf{g}_n = \nabla R(\mathbf{x}_n). \tag{4}$$

Because it depends only on $\mathbf{x}_n$, one can perform regularization in parallel with back projection and in any voxel order.

Finally, the image is updated as follows:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{D}(\mathbf{b}_n + \mathbf{g}_n), \tag{5}$$

where $\mathbf{D}$ denotes a diagonal matrix that is precomputed prior to iterating using optimization transfer principles [8]. We also enforce non-negativity in this step.

Parallelism is readily available in each of these steps; in principle, one could launch individual execution threads to calculate each element in $\mathbf{x}_n$, $\mathbf{r}_n$, $\mathbf{b}_n$, and $\mathbf{g}_n$. In practice, it is more efficient to group calculations into threads that allow common sub-expressions to be factored out of inner-most loops. Most existing statistical reconstruction implementations [4, 8] use programming interfaces such as POSIX threads, which allow concurrent operation on a single copy of $\mathbf{x}_n$, $\mathbf{r}_n$, $\mathbf{b}_n$, and $\mathbf{g}_n$ stored in a shared main memory. As each thread updates a disjoint subset of each matrix, the threads can proceed without synchronization, except for a global barrier between each step.

Though shared main memory provides a simple abstraction, it limits performance scalability. Far greater performance can be achieved by scaling a workload to execute on a large cluster. Distributed computation in a cluster is particularly cost-effective in cloud computing environments, where clusters can be time-shared and compute time leased by the hour at low cost.

B. Parallelizing over a cluster

A large cluster can bring far more compute cores to bear on a problem, but data updates must be transmitted explicitly between compute nodes. The central challenge of implementing statistical reconstruction on such a cluster lies in orchestrating this communication, requiring a fundamentally different implementation approach than conventional parallel statistical reconstruction.

Unfortunately, data structures are not easily partitioned and distributed among nodes in statistical reconstruction algorithms for X-ray CT. Both forward and back projection require essentially all data in a particular transaxial slice. Data can be somewhat partitioned along the scan axis, particularly for small cone angles that limit the axial interaction distance between image data and views. For a helical CT chest scan with 320 slices, we partition the image volume into, say, 5 "slabs" of 64 slices each and reconstruct each of those slabs by an independent set of compute nodes. Because of the "long object problem" in helical CT, to reconstruct 64 slices of interest, we also reconstruct 32 padding slices on each end of the slab. These padding slices are discarded after reconstruction, and thus represent a somewhat undesirable overhead of the slab-partition approach. This overhead means that it is inefficient to partition the volume into many more slabs with fewer slices per slab.

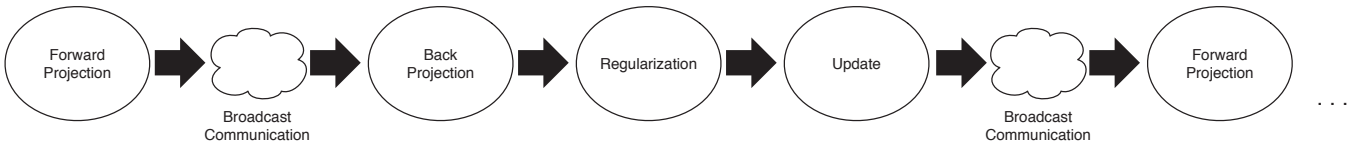


Fig. 1: Visual representation of computation phases. Arrows represent global barriers between steps.

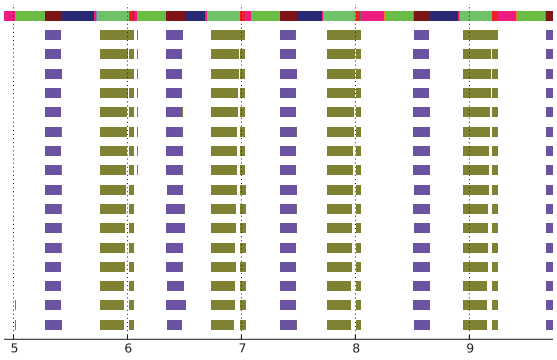


Fig. 2: Sample execution timeline illustrating phases of computation and communication. The color of Line 1 indicates the current algorithmic step. Dark red indicates forward projection, dark blue indicates communication after forward projection, light green indicates back projection, light red indicates regularization, purple indicates update, neon green indicates communication after the update step, and pink indicates barrier synchronization. The remaining lines indicate forward progress of individual threads on each core.

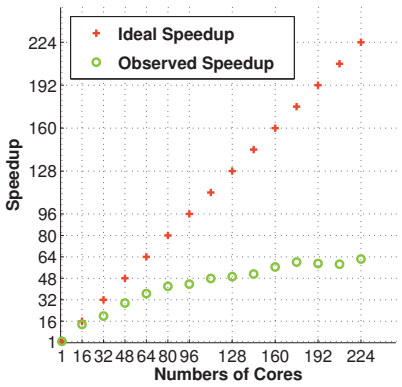


Fig. 3: Speedup of a single iteration of one slab for varying node configurations.

After each algorithm step, image/sinogram updates are broadcast and merged with results from all other nodes participating in the slab computation. We use the Message Passing Interface (MPI) for inter-node communication. MPI provides a means of sending and receiving data both synchronously (blocking) and asynchronously (non-blocking), as well as creating global barriers that prevent any node from proceeding past the point of the barrier until all nodes have reached it.

A straightforward communication approach places a significant burden on the interconnection network between servers. At the end of any given step, updating each node's copy of $\mathbf{x}$ requires each node to broadcast a copy of its portion of the data to $N-1$ other nodes, where N is the number of nodes participating in a given slab's reconstruction. An entire X-ray CT image volume for a helical scan can occupy about a gigabyte of memory (in single float precision). With a network bandwidth of 10 gigabits per second (as in Amazon's EC2 system), transferring several copies of the entire image volume incurs considerable delay. Communication is needed multiple times in each subset and iteration, so the total time spent sending and receiving over the network ultimately limits performance scalability. Hence, optimizing communication is critical. Partitioning the problem into smaller-sized slabs is a first step towards reducing communication. We also broadcast only those arrays that must be synchronized, which are the residual $\mathbf{r}_n$ and the updated image $\mathbf{x}_{n+1}$.

Figure 1 illustrates the reconstruction steps for a single subset. The arrows represent global barriers between each step, and the clouds represent the two necessary broadcast steps.

Figure 2 illustrates the computation timeline over about four (out of 12) subset updates, derived from measured results for execution on 10 nodes of 16 cores each. Each colored segment of the top bar represents an individual task, and the segment length indicates the amount of time taken for that task (in seconds). The 16 bottom bars show execution time for each core in the first node to perform the task identified by the color in the topmost bar. Synchronization and communication time (dark blue, pink, and neon green in the top bar; large blank regions in the remaining bars) occupy a significant fraction of overall execution time.

III. Experimental Results

We report on our distributed version of PWLS-OS implemented in the C99 programming language using the openMPI and POSIX thread libraries. Our implementation produces identical output to that of a previously existing multithreaded version, confirming our methods do not sacrifice image quality.

Our test environment consisted of Amazon EC2 HPC cc2.8xlarge nodes, each having dual eight-core 2.6 GHz Xeon processors, 60.5 GB of memory, and 10 gigabit ethernet. We used Amazon's group placement policy for all experiments to ensure nodes were located physically close to each other.

We used simulated helical CT data where the image volume for a single slab is $512 \times 512 \times 128$ slices with a 70 cm transaxial field of view (FOV) and 0.625 mm slice thickness. Out of an entire 9-turn helical scan, with pitch = 63/64, we used 3 turns (2952 views) for reconstructing each slab. The views were each 64 rows by 888 channels. We use the separable footprint projector [9]. Voxels within the 70 cm FOV are reconstructed using [8]. We used 12 subsets because that is suitable for our latest accelerated OS algorithm [2].

Figure 3 shows the computational speedup for a single iteration plotted for a single core up to 224 cores (each node contains 16 cores) for our implementation compared to ideal (linear) speedup. Speedup [11] for a system with n cores is

$$\frac{\text{Time taken for 1 core}}{\text{Time taken for } n \text{ cores}}. \tag{6}$$

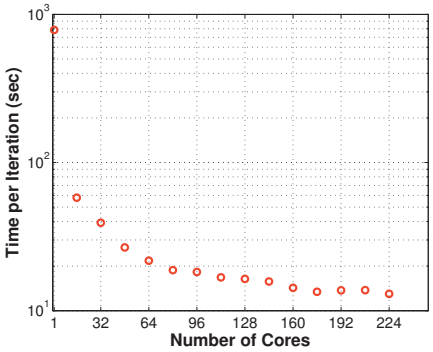


Fig. 4: Timing results for a single iteration on one slab for varying node configurations.

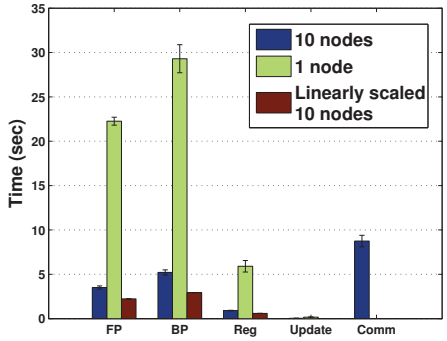


Fig. 5: Time (per iteration) spent in each step for one slab in the 10-node and 1-node cases compared to ideal linear scaling.

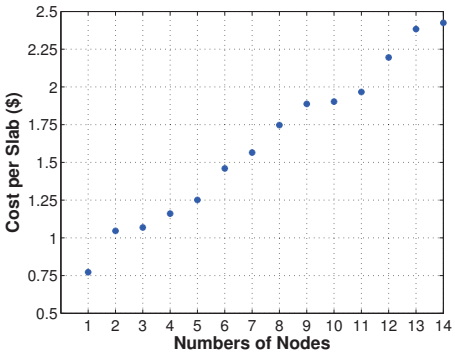


Fig. 6: Cost for a single slab of a 20-iteration scan for varying node configurations.

Figure 3 shows that the observed speedup growth slows as nodes are added due to communication time and bandwidth constraints. This behavior results in a knee in the speedup curve as it approaches a maximum speedup of about 64.

Figure 4 depicts the time taken to complete a single iteration for 1 core to 224 cores in 1-node increments. As expected, the time curve decreases more and more slowly as the number of nodes increases, asymptotically approaching a minimum of about 12 seconds. Communication begins to dominate around 160 cores (10 nodes), at which point the benefit of using more nodes becomes insignificant.

Figure 5 shows computation times of individual steps for the observed 10-node, 1-node, and ideal linear scaling 10-node cases. Linear scaling for n nodes is defined as achieving a speedup of n, so a linearly scaled 10 node system would take one tenth of the time of the 1 node case per iteration. The actual time taken in each compute step (blue) is reasonably close to the ideal linearly scaling case (red). Thus individual steps scale well even though total iteration time exhibits much less than linear speedup due to communication time. The significance of communication time is evident; it accounts for nearly half the time per iteration in the 10-node case.

Figure 6 plots the cost of running 20 PWLS-OS iterations to reconstruct a single 128-slice slab versus the number of nodes used. The cost increases monotonically because communication is such a significant factor in performance. The total cost, however, is still inexpensive.

Based on diminishing returns when using more than 10 nodes, we focus on the 10-node case as a reasonable configuration. Using 10 nodes (160 cores) per slab, the total time for reconstructing a 5-slab scan (320 usable slices) is $\frac{15 \text{ sec}}{\text{iteration}} \times 20 \text{ iterations} = 300$ seconds. Likewise, the cost of performing a reconstruction is $\frac{10 \text{ nodes}}{\text{slab}} \times 5 \text{ slabs} \times \frac{\$0.00067}{\text{sec}} \times \frac{15 \text{ sec}}{\text{iteration}} \times 20 \text{ iterations} = \$10$. For a longer helical scan with 640 slices the cost would scale to \$20 (by using more nodes for the additional slabs) but the 5-minute reconstruction time would remain unchanged.

IV. Summary and Conclusions

Use of statistical reconstruction methods is impeded by their computation time. We have investigated using commercial cloud computing to improve the speed of MBIR through parallel computing. Our results, generated using Amazon's EC2 service, show that even with significant communication overhead, attractive reconstruction times (5 minutes) can be achieved at a low price (\$10). If high resolution targeting of a region of interest (ROI) is needed, then a two-stage reconstruction will be needed [12] that would increase the time and cost accordingly. As expected, node-to-node communication is a limiting factor on performance, even for a relatively small number of nodes. Future work includes reducing communication by using data compression techniques, and by devising iterative algorithms that do not require full synchronization after every update.

References

[1] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.

[2] D. Kim, S. Ramani, and J. A. Fessler. Ordered subsets with momentum for accelerated X-ray CT image reconstruction. In *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2013. To appear.

[3] D. Kim and J. A. Fessler. Parallelizable algorithms for X-ray CT image reconstruction with spatially non-uniform updates. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 33–6, 2012.

[4] M. Wu and J. A. Fessler. GPU acceleration of 3D forward and backward projection using separable footprints for X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 56–9, 2011.

[5] B. Meng, G. Pratx, and L. Xing. Ultrafast and scalable cone-beam CT reconstruction using MapReduce in a cloud computing environment. *Med. Phys.*, 38(12):6603–9, December 2011.

[6] S. Srivastava, A. R. Rao, and V. Sheinin. Accelerating statistical image reconstruction algorithms for fan-beam x-ray CT using cloud computing. In *Proc. SPIE 7961 Medical Imaging 2011: Phys. Med. Im.*, page 796134, 2011.

[7] J. Gregor. Distributed multi-core implementation of SIRT with vectorized matrix kernel for micro-CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 64–7, 2011.

[8] D. Kim, D. Pal, J-B. Thibault, and J. A. Fessler. Improved ordered subsets algorithm for 3D X-ray CT image reconstruction. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 378–81, 2012.

[9] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29(11):1839–50, November 2010.

[10] R. C. Fair. On the robust estimation of econometric models. *Ann. Econ. Social Measurement*, 2:667–77, October 1974.

[11] J. L. Hennessy and D. A. Patterson. *Computer architecture: A quantitative approach*. Morgan Kaufmann, San Francisco, 4 edition, 2006.

[12] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. In *Proc. SPIE 6142 Medical Imaging 2006: Phys. Med. Im.*, page 614223, 2006.

Extended Volume Image Reconstruction Using the Ellipse-Line-Ellipse Trajectory for a C-arm System

Zhicong Yu, Frédéric Noo, Günter Lauritsch, and Joachim Hornegger

***Abstract*—Recently, we proposed the Ellipse-Line-Ellipse trajectory for extended volume imaging with a C-arm system. Knowledge of the R-line coverage of this trajectory is well understood, but how to use the R-lines remains unclear. In this work, we establish a scheme for efficient and practical usage of the R-lines of this trajectory. Using computer-simulated data, we demonstrate this scheme by reconstruction results from the differentiated backprojection method.**

I. Introduction

We are interested in the development of extended volume cone beam (CB) computed tomography (CT) using the C-arm system in interventional radiology. For some intraoperative or emergency cases where the entire aorta or the spine needs to be examined, this tool could be crucial for patient health, particularly because it would prevent transferring the patient to the CT room, which is time consuming and increases risk to the patient.

Currently, C-arm systems employ a circular trajectory, which does not satisfy Tuy's condition and can not address extended volume imaging. A more sophisticated trajectory is needed. The helical trajectory has been successful in traditional CT systems, but it is not feasible on a C-arm system. Due to the open design and lack of slip-ring technology, a C-arm can not rotate infinitely in a single direction; in most cases, it can only allow a short scan. To overcome this mechanical limitation, the reverse helix [1] was proposed, which is like the helix, but reverses its rotational direction after a certain angular length. However, this trajectory does not have sufficient R-line coverage (An R-line is any segment of line that connects two source positions of the trajectory) in the middle of its convex hull, and thus theoretically-exact and stable (TES) reconstruction is challenging. Another possible solution would be the Arc-Line-Arc trajectory [2]. Nevertheless, this trajectory requires a full scan to achieve full R-line coverage in the region-of-interest (ROI), regardless of the size of the ROI, which is impractical for a C-arm system in most cases.

Recently, we have proposed a new geometry called the Ellipse-Line-Ellipse (ELE) trajectory [3], [4]. This trajectory possesses a reverse pattern in its rotational direction, and thus is feasible on a C-arm system. More importantly, we have proved that, through a simple configuration using a short scan, this trajectory provides sufficient R-lines for a typical cylindrical ROI that is centered on the patient table and inside the convex hull of the ELE trajectory.

At this stage, the R-line coverage of the ELE trajectory is well understood according to [3], [4], however, an efficient and practical scheme to select R-lines for TES image reconstruction yet needs to be established. In this work, we provide such a scheme for each point within the ROI such that the usage of the projection data varies smoothly when the point of interest in ROI moves continuously. We demonstrate this scheme by computer simulations using the method of differentiated backprojection followed by inverse Hilbert transform (DBP-HT) [5]. Reconstruction results show good image quality with smooth transition in transversal, coronal and sagittal directions.

Zhicong Yu and Frederic Noo are with Department of Radiology, University of Utah, Salt Lake City, UT, USA; Günter Lauritsch is with Siemens AG, Healthcare Sector, Forchheim, Germany; Zhicong Yu and Joachim Hornegger are with The Chair of Pattern Recognition, University of Erlangen-Nuremberg, Erlangen, Germany. This work was partially supported by a grant of Siemens AG, Healthcare Sector and by the U.S. National Institutes of Health (NIH) under grant R21 EB009168 and R01 EB007236. The concepts presented in this paper are based on research and are not commercially available. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

II. The ELE trajectory

A. Geometry

The ELE trajectory lies on a cylindrical surface of radius R that is centered on the z-axis. This trajectory consists of two elliptical arcs connected by a segment of line as shown in the left of Figure 1, and we call the three components as the upper T-arc, T-line and the lower T-arc, respectively. In the attached (x, y, z)-coordinate system, the upper and lower T-arcs are mirror symmetric relative to the (x, y)-plane, and they lie in planes that are perpendicular to the (x, z)-plane, whereas the T-line is parallel to the z-axis; see the middle and right of Figure 1. Note that the ELE trajectory can be periodically repeated along the z-direction, and thus is suitable for extended volume imaging.

Let λ be the polar angle, and let γ_m be the fan-angle. We denote $\underline{a}_u(\lambda)$ and $\underline{a}_l(\lambda)$ as vertex points on the upper and lower T-arcs, respectively, and refer to $\underline{b}(z)$ as a vertex point on the T-line. By definition, we have

$$\begin{aligned}\underline{a}_u(\lambda) &= (R\cos\lambda, R\sin\lambda, H + \Delta H\cos\lambda),\\ \underline{b}(z) &= (R\cos\gamma_m, -R\sin\gamma_m, z),\\ \underline{a}_l(\lambda) &= (R\cos\lambda, R\sin\lambda, -H - \Delta H\cos\lambda),\end{aligned}$$

where $\lambda \in [-\gamma_m, \pi + \gamma_m]$ and $z \in [-H - \Delta H\cos\gamma_m, H + \Delta H\cos\gamma_m]$, with H and ΔH being configuration parameters as depicted in the right of Figure 1.

B. R-line coverage and ROI design

The R-line coverage of the ELE trajectory is composed of three parts, i.e., R-lines that are generated by connecting the

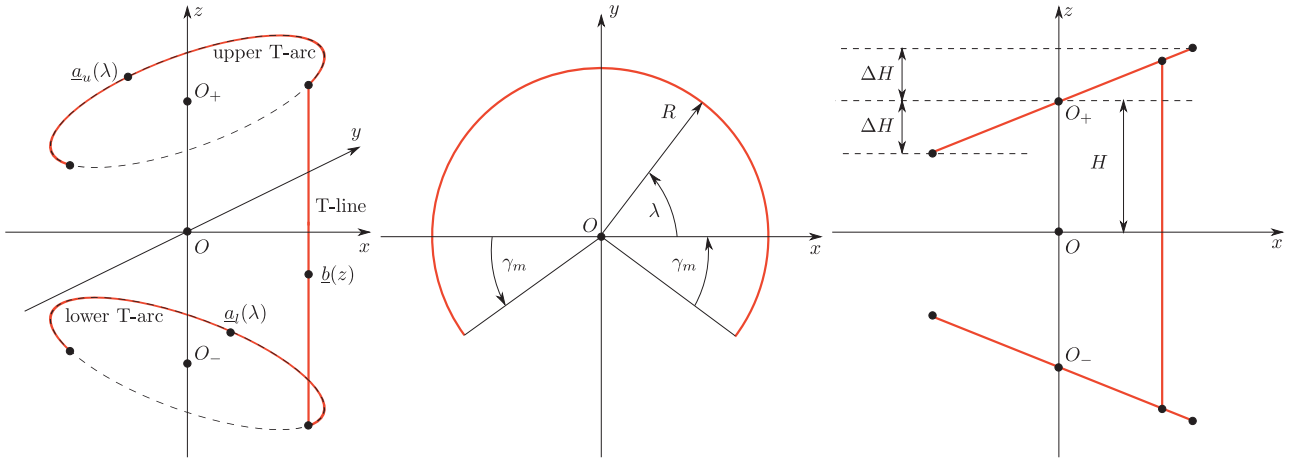


Fig. 1. Left: 3D illustration of the ELE trajectory. Middle: orthogonal projection onto the (x, y)-plane. Right: orthogonal projection onto the (x, z)-plane.

upper T-arc and the T-line (AL), the T-line and the lower T-arc (LA), as well as the upper and lower T-arcs (AA). Because the ELE trajectory is mirror symmetric relative to the (x, y)-plane, its R-line coverage is mirror symmetric as well. Hereafter, we will only focus on the R-line coverage for the space that is above the (x, y)-plane and within the convex hull of the ELE trajectory.

According to [4], for a given cylindrical ROI of radius $r < 0.85R$, full R-line coverage in the ROI can be guaranteed using the following configuration:

$$\gamma_m = \arcsin(r/R) \qquad \text{and} \qquad \Delta H/H = r/R.$$

C. Selection of R-lines

We define the surface generated by connecting one point on the T-line or the lower T-arc to all the points on the upper T-arc as the R-line surface, as shown in the left of Figure 2. When the convergent point of the R-line surface moves from the top point to the bottom point along the T-line, we obtain the blue R-line surfaces; whereas when the convergent point continues to move along the lower T-arc, we get the green R-line surfaces. We call the region covered by the R-line surfaces from the AL trajectory as the blue region, and refer to the region covered by the R-line surfaces from the AA trajectory as the green region.

It can be shown that when the convergent point on the lower T-arc reaches a certain polar position λ_c, every point in the ROI above the (x, y)-plane is covered by R-lines. For a typical ROI with radius $r < 0.5\,R$, $\lambda_c = 0$. Therefore, besides the upper T-arc and the T-line, only a small portion of the lower T-arc is needed for TES reconstruction in the ROI that is above the (x, y)-plane. In this work, for TES image reconstruction of the ROI above the (x, y)-plane, the R-lines forming the blue and green regions are selected.

The R-lines from the AL trajectory (blue) can cover a large portion of the ROI. However, in the first and second quadrants, some regions of the ROI are only covered by the R-lines from the AA trajectory, as shown in the middle of Figure 2. Let Q be a point in the green region, and refer to $\mathcal{L}(Q)$ as the line parallel to the z-axis that goes through Q, as shown in the right of Figure 2. As illustrated, the upper part (AB) of $\mathcal{L}(Q)$ is covered by the R-lines from the AL trajectory and the lower part (BC) is covered by the R-lines from the AA trajectory. Note that, C is not necessary the lowest point of the R-line coverage along $\mathcal{L}(Q)$ when the source position on the lower T-arc moves from $-\gamma_m$ to λ_c.

Some of the blue region is not only covered by R-lines from the AL trajectory, but also by the R-lines from the AA and LA trajectories. We disregard these contributions from the AA and LA trajectories so as to create a continuous scan and reconstruction flow. To the same reason, some of the green region is covered by two R-lines from the AA trajectory, but we only choose the one that has $\lambda \in [-\gamma_m, \lambda_c]$.

III. Reconstruction Method

Using the concept of an R-line surface, we perform the TES reconstruction using the DBP-HT method [5]. Let $\underline{x}$ be a 3D point, and $f(\underline{x})$ be the attenuation coefficient of the object at $\underline{x}$. We assume that $f(\underline{x})$ is compactly supported.

To perform the DBP-HT method in an efficient way, we now introduce a new coordinate system to describe the R-line surfaces, as shown in Figure 3. Let $\underline{b}(h)$ be a point on the T-line, and $\underline{a}_u(\lambda_*)$ be a point on the upper T-arc. We denote a point on the R-line that connects $\underline{b}(h)$ and $\underline{a}_u(\lambda_*)$ as $\underline{s}(h, \gamma, t)$. Let $\underline{\alpha}$ be the vector pointing from $\underline{b}(h)$ to $\underline{s}(h, \gamma, t)$, and refer to $\underline{\alpha}_{xy}$ as the orthogonal projection of $\underline{\alpha}$ onto the (x, y)-plane. Then γ and t are the polar angle and magnitude of $\underline{\alpha}_{xy}$, respectively. Because we are only interested in reconstructing the ROI, we have $\gamma \in [\pi - 2\gamma_m, \pi]$ and $t \in [R - r, R + r]$.

The coordinate system for describing the R-lines from the AA trajectory is similar to that from the AL trajectory, except that we use ω to indicate a point on the lower T-arc, i.e., $\underline{a}_l(\omega)$. For $r < 0.5R$, $\omega \in [-\gamma_m, 0]$. In this case, the range of γ is dependent on ω, i.e., $\gamma \in [\pi - \gamma_m + w, \pi + 0.5\gamma_m + 0.5\omega]$. The lower bound of γ is set to make sure that the ROI is covered by the R-line surfaces, whereas the upper bound of γ is set to make sure that the R-line surfaces do not include R-lines beyond the endpoint of the upper T-arc.

Based upon the above description, image reconstruction from the ELE trajectory using the DBP-HT method can be

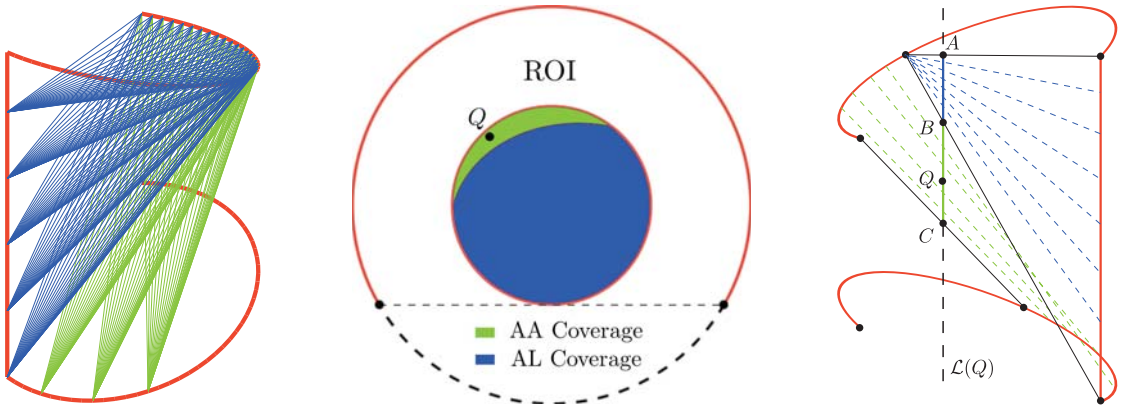


Fig. 2. Selection of the R-lines of the ELE trajectory for TES image reconstruction. Left: the R-line surfaces associated to the AL (blue) and AA (green) trajectories. Middle: the R-line coverage in the ROI in the (x, y)-plane. Right: R-line coverage along $\mathcal{L}(Q)$ that is from the AL (blue, composed of the upper T-arc and the T-line) and AA (green) trajectories.

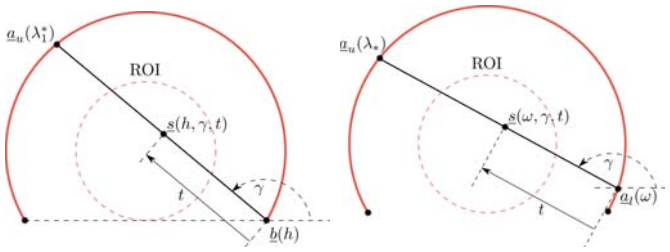


Fig. 3. The R-line surfaces of from the AL (left) and AA (right) trajectories can be described by (h, γ, t)- and (ω, γ, t)-coordinate systems, respectively.

achieved by the following steps.

- perform the view-dependent differentiation for all the projection data of the ELE trajectory;
- for each (h, γ, t) or (ω, γ, t), perform backprojection using the R-line segments that are selected in Section II-C;
- perform inverse Hilbert transform along t;
- rebin the reconstruction from the (h, γ, t)- or (ω, γ, t)-coordinate system to the (x, y, z)-coordinate system.

Among the four reconstruction steps, the rebin process deserves particular attention. For a given $\underline{x}$, if there is an R-line through this point from the AL trajectory, we should perform the rebinning from the (h, γ, t)-coordinate system to the (x, y, z)-coordinate system. Otherwise, we should perform the rebinning from the (ω, γ, t)-coordinate system to the (x, y, z)-coordinate system. The rebinning for the latter case is not trivial, and a detailed solution can be found in [4].

IV. Results

In this section, we present our numerical results using computer-simulated data. A modified FORBILD head phantom was adopted for data simulation. This phantom was obtained by stretching most of the FORBILD head phantom in the x-direction by a factor of 1.25, so that the horizontal slices of this phantom is circular, as shown in Figure 4. This modification was made for a better illustration of the reconstruction results from the AA trajectory.

The largest slice of the modified phantom appears at $z = 0$ cm with radius 12 cm, which, therefore, was set as the radius

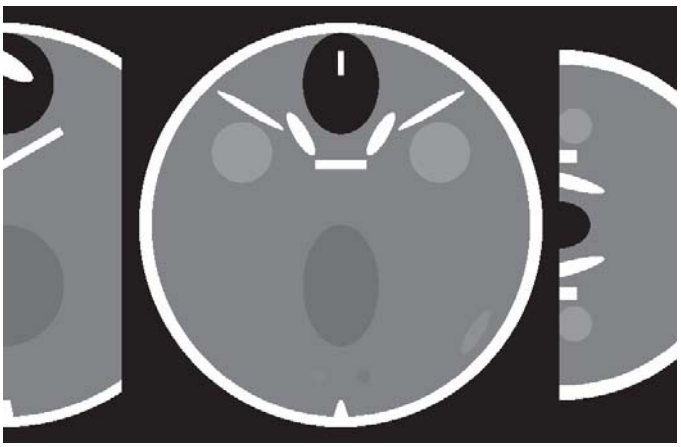


Fig. 4. The modified FORBILD head phantom. Left: sagittal slice at $x = 0$ cm with $z \in [0, 7]$ cm. Middle: transversal view at $z = 0$ cm. Right: coronal view at $y = 6$ cm with $z \in [0, 7]$ cm. Display window: $[0, 100]$ HU.

of the ROI. For data acquisition, a flat panel detector of bin size $0.06\,\text{cm} \times 0.06\,\text{cm}$ was used. This detector was large enough to avoid truncation. The scan radius was 30 cm, and the source-to-detector distance was 45 cm.

Regarding the ELE trajectory, we chose $H = 5$ cm and $\Delta H = 2$ cm so that R-line coverage was sufficient in the ROI. For each of the upper and lower T-arcs, 500 CB projections were generated over the angular range $[-24°, 204°]$, whereas for the T-line, 51 CB projections were generated. Note that the z-range of the T-line was automatically generated using H, ΔH and the angular range of the T-arcs. All CB projections were acquired using quarter detector pixel shift.

Regarding the DBP-HT reconstruction algorithm, we implemented the view-dependent differentiation according the the scheme presented in [6]. The resolution control parameter ε used in our experiment was 0.001. The inverse Hilbert transform was implemented according to Equation 18 of [7], and the constant C_t in this equation was calculated using Equation 20 from the same paper.

Reconstructions on the R-line surfaces from the AL and AA trajectories are shown in Figure 5. For the AL trajectory,

we selected 200 R-line surfaces with their convergent points evenly distributed along the T-line. For each AL R-line surface, 900 points were sampled evenly for γ over the range $[132°, 180°]$, whereas 900 points were sampled for t over the range $[16.5\,\text{cm}, 33.5\,\text{cm}]$. For the AA trajectory, we selected 225 R-line surfaces with their convergent points uniformly distributed along the lower T-arc over the range $[-24°, 0°]$. For each w, we evenly sampled 900 points for γ over the range $[\pi - \gamma_m + w, \pi + \gamma_m + w]$, with $\gamma_m = 24\pi/180$. The grid for t was the same as that of the R-line surfaces from the AL trajectory.

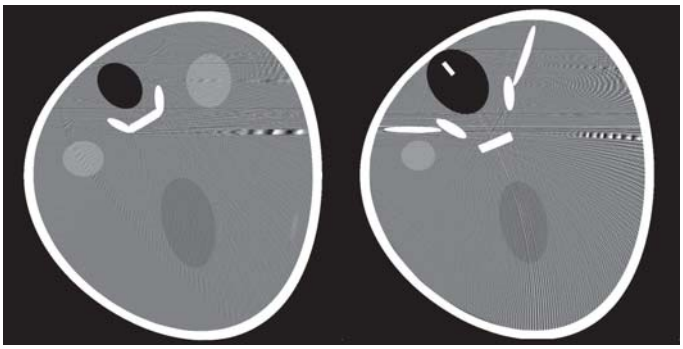


Fig. 5. Left: the reconstruction on the R-line surface at $h = 0.0343\,\text{cm}$. Right: the reconstruction on the R-line surface at $w = -22.4°$. Display window: $[0, 100]$ HU.

We obtained the final image reconstruction through rebinning using triple linear interpolation with isotropical voxels of size $0.02\,\text{cm}\times0.02\,\text{cm}\times0.02\,\text{cm}$. Part of the final image was rebinned from the AL trajectory, and the rest was rebinned from the AA trajectory. The rebinned images at $z = 0\,\text{cm}$ from the AL and AA trajectories are shown in Figure 6. The final rebinned results by using both AA and AL trajectories are shown in Figure 7. As illustrated, the final reconstruction results have good image quality with smooth transition in coronal, sagittal and transversal directions.

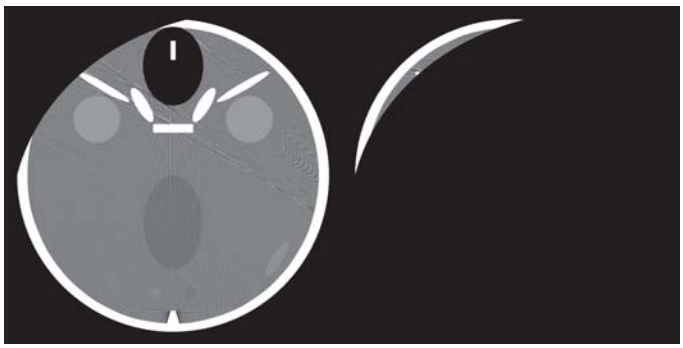


Fig. 6. Left: image rebinned from the R-line surfaces of the AL trajectory. Right: image rebinned from the R-line surfaces of the AA trajectory. Image position: $z = 0\,\text{cm}$. Display window: $[0, 100]$ HU.

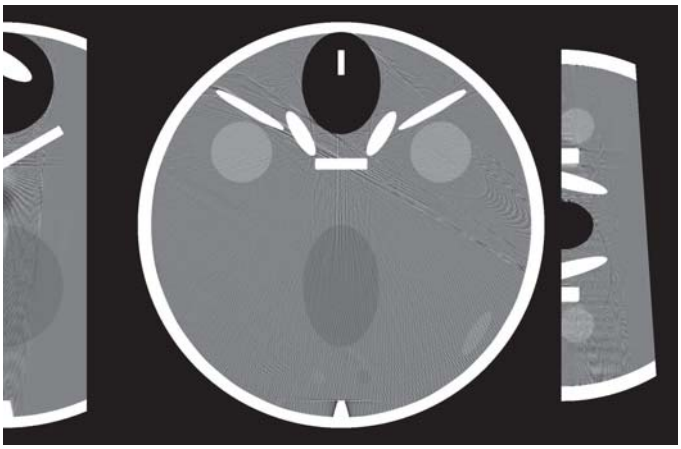


Fig. 7. Final image reconstruction rebinned through triple linear interpolation using isotropical voxels of size $0.02\,\text{cm}\times0.02\,\text{cm}\times0.02\,\text{cm}$. Top: sagittal view at $x = 0\,\text{cm}$. Middle: transversal view at $z = 0$. Bottom: coronal view at $y = 6\,\text{cm}$. Display window: $[0, 100]$ HU.

V. Conclusion and discussion

We have proposed an efficient and practical method to use the R-lines of the ELE trajectory for TES image reconstruction. For reconstruction of the ROI above the (x, y)-plane, besides the upper T-arc and the T-line, only a small continuous portion of the lower T-arc that is connected to the T-line is needed.

Although we only demonstrated how to address the R-lines for the ROI above the (x, y)-plane, it is straight forward to extend this method to the whole ELE trajectory as well as its duplicates. Also note that, in this work, we have used the DBP-HT method for image reconstruction, however, other TES methods such as Katsevich-type algorithms are also applicable.

References

[1] S. Cho, D. Xia, C. A. Pellizzari, and X. Pan, "A BPF-FBP tandem algorithm for image reconstruction in reverse helical cone-beam CT," *Med. Phys.*, vol. 37, no. 1, pp. 32–39, Jan 2010.

[2] Z. Yu, A. Wunderlich, F. Dennerlein, G. Lauritsch, and F. Noo, "Line plus arc source trajectories and their r-line coverage for long-object cone-beam imaging with a c-arm system," *Phys. Med. Biol.*, vol. 56, no. 12, p. 3447, 2011.

[3] Z. Yu, F. Noo, G. Lauritsch, F. Dennerlein, and J. Hornegger, "Ellipse-line-ellipse source trajectory and its r-line coverage for long-object cone-beam imaging with a c-arm system," in *SPIE Conference Series*, vol. 8313, 2012, p. 117.

[4] Z. Yu, F. Noo, Y. Mao, F. Dennerlein, G. Lauritsch, and J. Hornegger, "Ellipse-line-ellipse source trajectory and its r-line coverage for extended volume imaging using a c-arm system," *In preparation for submission to Phys. Med. Biol.*

[5] H. Schöndube, *Helical Cone-Beam Computed Tomography using the Differentiated Backprojection*. Shaker Verlag, Aachen, Germany, 2010.

[6] F. Noo, S. Hoppe, F. Dennerlein, G. Lauritsch, and J. Hornegger, "A new scheme for view-dependent data differentiation in fan-beam and cone-beam computed tomography," *Phys. Med. Biol.*, vol. 52, no. 17, p. 5393, 2007.

[7] F. Noo, R. Clackdoyle, and J. Pack, "A two-step hilbert transform method for 2d image reconstruction," *Phys. Med. Biol.*, vol. 49, no. 17, p. 3903, 2004.

Image Reconstruction from an Axial Short Scan using a Katsevich type algorithm

Considerations for Repeatable Quantitative Imaging

Brian E. Nett[1], Kai Zeng[2], and Jed D. Pack[3]
[1] GE Healthcare, 3000 N. Grandview Blvd., W-1180, Waukesha WI USA
[2] Formerly with GE Global Research Center, One Research Circle, Niskayuna, NY USA
[3] GE Global Research Center, One Research Circle, Niskayuna, NY USA

***Abstract*—In computed tomography reconstruction there are several scenarios where data is acquired with a large cone angle (e.g. microCT, breast CT, radiation therapy, surgical guidance). In many cases only a short-scan of data is available: on many C-arm interventional or surgical systems the mechanical constraints of the system limit the acquisition to a short-scan (π + full fan angle). The method frequently employed for these acquisitions is to use a standard FDK based reconstruction with an approximate view based weighting based on the 2D (fan-beam) view weighting. The aim of this work is to compare this standard reconstruction method with an approximate reconstruction algorithm derived using the Katsevich framework for exact image reconstruction. The assessment in this work employs a realistic anatomical phantom (the XCAT) and focuses on the possibility of performing quantitative imaging from only a short scan acquisition. The results demonstrate that the algorithm derived using the exact framework provides more quantitative results and the central view angle dependence of the acquisition is significantly reduced compared with the standard reconstruction method.**

I. Introduction

The goal of this work is to compare this standard reconstruction method with an approximate reconstruction algorithm derived using the Katsevich framework for exact image reconstruction [1], [2]. The derivation of this algorithm was previously published, Arc based Cone-beam reconstruction algorithm using an Equal weighting scheme (ACE), along with an initial evaluation using the Shepp-Logan phantom and a Defrise type phantom. The assessment in the current work employs a more realistic anatomical phantom and the evaluation is aimed at demonstrating improved quantitative imaging from only a short scan acquisition. In addition to assessing the reconstruction accuracy for a single acquisition we assess the variability in the reconstruction values as a function of the central view angle. For instance in C-arm imaging the reconstruction values of an object would change depending of the patient position, e.g. prone or supine. Additionally, for perfusion measurements it is important that images reconstructed at different time points do not vary in reconstruction value purely due to the difference in the central view angle.

In wide-cone axial computed tomography there are many factors which make quantitative imaging difficult including scatter, beam hardening, motion, data truncation and the effects of image reconstruction. In this work we focus exclusively on the effects of image reconstruction. In most evaluations of quantitative accuracy simplistic (non-anthropomorphic) phantoms are used; and typically if data from a system is used it is difficult to separate the contribution of the inaccuracy related to the different components. We use simulation data in this case to separate the effects of the image reconstruction from the other physical non-idealities such as scatter and beam hardening.

Since the axial trajectory is easy to achieve mechanically it is very popular for a number of medical systems which collect projection data for CT reconstruction. While many other options have been proposed for trajectories which offer exact reconstruction, the ubiquitous nature of the axial trajectory makes it an important case for algorithms with improved reconstruction accuracy, even if exact reconstruction is not achievable.

II. Methods

A. Simulations

The simulations were performed with the standard male phantom of the XCAT NURBs based phantom, where the forward projection values were generated using the CatSim software package. A cylindrical detector was simulated with a fan angle of $48°$, a cone angle of $14.6°$, 888 detector channels.. In the simulations multiple sampling rays were used for each measurement. To assess the angular dependence of the reconstruction accuracy short scans were performed with $15°$ spacing covering $360°$. The simulations were monoenergetic at 70 keV, and a beam current of 550 mA was used for all the simulations shown in the figures and no quantum noise was simulated for the data used in the plots of quantitative metrics.

B. Implementation

The derivation based on the Katsevich framework and initial implementation of the ACE algorithm has been previously published [3], therefore here we only provide a high level summary and *highlight* new features in the algorithm implementation presented here.

- *Input data is from an equi-angular detector in the 3rd generation geometry.*
- Differentiate (along row, column, view) and scalar weight data as described in [4].

- Perform the Hilbert Filtering for three groups, for the two non-horizontal groups of filtering lines.
  - *Rebin from the cylindrical detector to a virtual flat panel detector*
  - *2D Bicubic interpolation onto the filtering directions on the virtual flat panel detector*
  - Hilbert Filter as described previously[3], *using a boosted kernel to match the spatial resolution of the FDK reconstruction (emperically matched MTF values by using the frequency boost* $\frac{MTF_{FDK}}{MTF_{ACE}}$*) and upsample by a factor of 2 compared with FDK prior to filter*
  - *2D Bicubic interpolation from the filtering directions onto virtual flat panel detector*
  - *Rebin from the virtual flat panel detector to the cylindrical detector*
- In the backprojection process, calculate the weight based on the (x-y) position of the voxels, and perform a weighted sum from the three filtering groups (weights are saved as volume typically aligns with z).

The short scan FDK implementation used for comparison here uses a Parker type weighting function [5] as in the comparison in the early ACE paper [3], and the cone-parallel geometry was used for the filtering and backprojection process.

C. Quantitative Evaluation

Images were reconstructed using a centered image volume (in-plane) with a field of view of 320mm, with an image grid on 512×512, where the slices were reconstructed with equal spacing covering a cone angle of $14.6°$. Given that only a short scan of data is available there are some portions of a cylindrical volume which have very little angular coverage due to z-truncation. In this study a mask has been applied such that only voxels which project onto the detector for all views are included in the analysis and image display. Since a numerical phantom has been used for this study it is possible to compare the quantitative image values with the ground truth phantom values. The reconstruction accuracy may vary based on the phantom composition so three automated Regions-of-Interest (ROIs) were defined on a slice by slice basis to quantify the accuracy and reproducibility of each reconstruction algorithm. In Figure 1 a single central slice is shown with the myocardium, fat and lung masks which are determined by the values in the phantom and then eroded with a disk structuring element of radius 5 pixels to prevent partial volume effect from influencing the results. The standard deviation of the voxel values as a funtion of central view angle is also computed as a metric for reconstruction reproducibility under different scanning conditions.

III. Results

Comparisons will be made between the Parker type weighted FDK (P-FDK) reconstructions and the ACE reconstruction via image comparisons and measures of quantitative accuracy and reproducibility with respect to view angle. In Figure 2 the central sagittal and coronal slices are compared

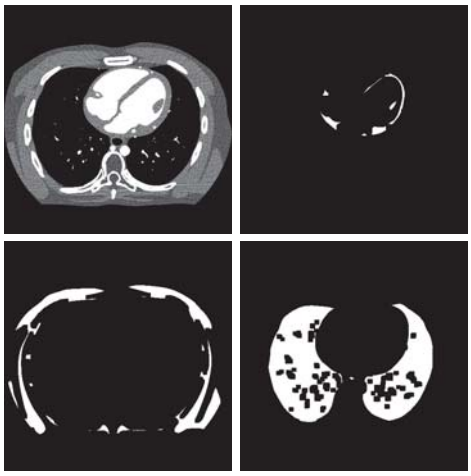


Fig. 1. Demonstration of the automated ROIs for a given slice. Upper left: Central slice used for demonstration purposes. Upper right: myocardium ROI. Lower left: fat ROI. Lower right: lung ROI.

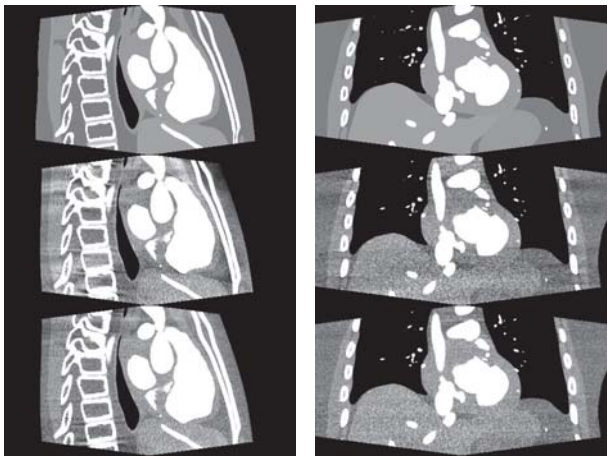


Fig. 2. Top: XCAT Phantom. Middle: P-FDK Reconstruction. Lower: ACE Reconstruction. Left: The central sagittal slice. Right: The central coronal slice. [-200 200] HU

between the P-FDK and ACE images. While some artifacts are still visibile due to the missing data the artifacts are greatly reduced in the ACE reconstructions. Sample axial slices are shown in Figure 3 where the reduction in artifacts is significant. The quantitative accuracy of the reconstructions for the three tissue types is shown in Figure 4, where both algorithms perform better in the central slices and the ACE values typically are closer to the true value.

The effect of changing the central view angle is demonstrated in the images shown in Figure 5, where the strong angular dependence of the P-FDK reconstruction is significantly reduced. The improvement in consistency across different central views is quantified by computing the standard deviation with respect to the central view angle of each voxel. Sample maps for the same slice are shown in Figure 6. Note that no low signal correction has been applied so there are some noise streaks which lead to some significant variance values

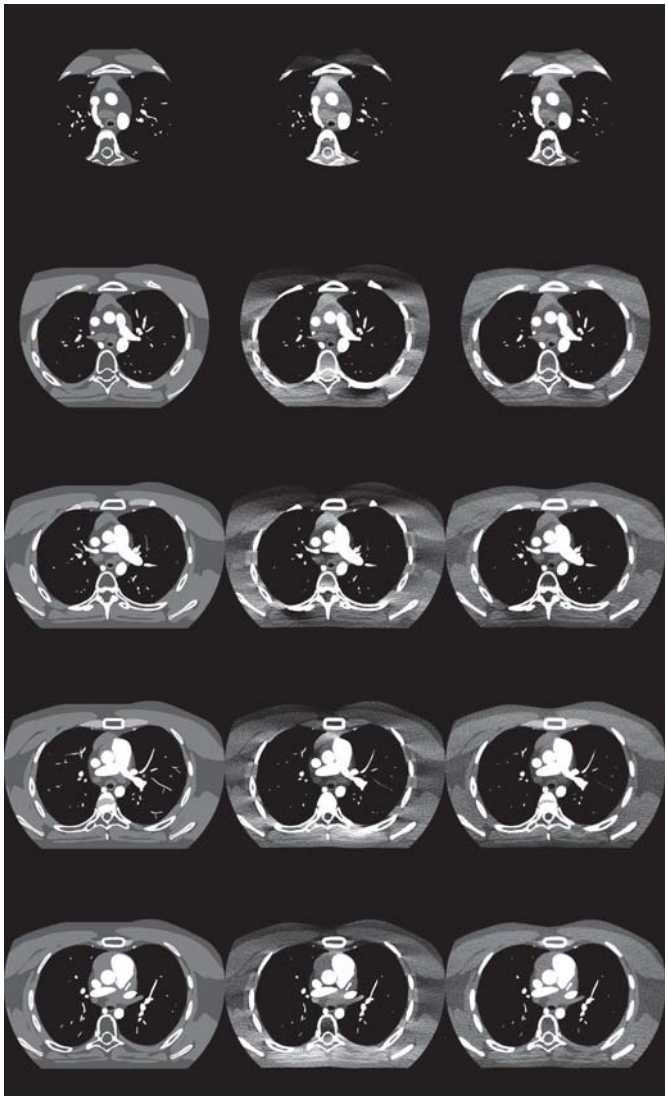


Fig. 3. Left: XCAT Phantom. Middle: P-FDK Reconstruction. Right: ACE Reconstruction. From the top to the bottom slices with cone angles of $6.27°$, $5.70°$, $5.13°$, $4.56°$, $4.0°$. [-200 200] HU

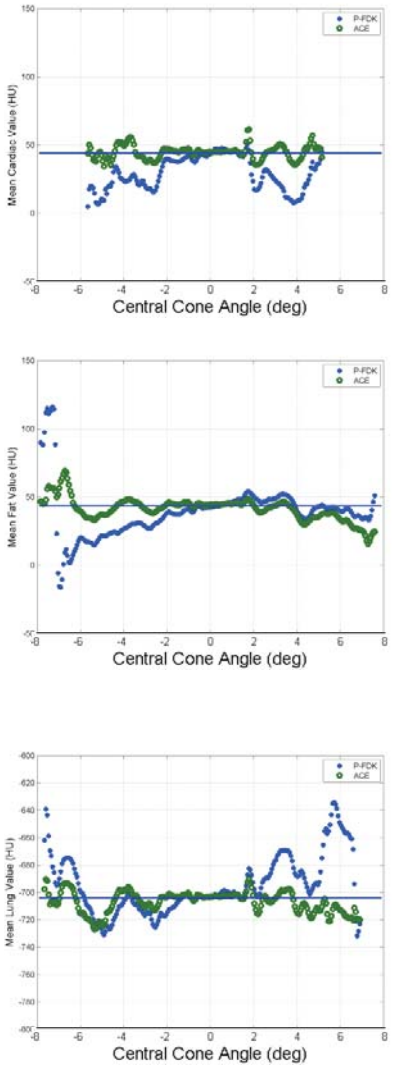


Fig. 4. Plots of the average CT number in the automated ROIs for cardiac, fat and lung tissue as a function of the z slice location. The blue line is the ground truth, the blue (x) are P-FDK and the green (o) are the ACE results.

in the ACE reconstructions but overall the standard deviations are much lower than the P-FDK results. The quantification of these results are given in Figure 7. The metric used here is the fraction of the automated ROI which has a standard deviation with respect to view angle of less than 10 HU. For these voxels one may reliably compare multiple acquisitions with different central view angles. As expected both algorithms performed well in the central slices and the z range of reproducible values was significantly larger for ACE than P-FDK.

IV. Conclusion

In this work the prospect of repeated quantitative CT image reconstruction was studied using a realistic anatomical phantom (ie the XCAT phantom). Multiple images volumes were reconstructed based on short scan data which have central view angles separated by $15°$. The standard FDK based reconstruction with a 2D Parker type weighting function (P-FDK) was compared with the ACE method. The ACE method was derived using the Katsevich framework for Filtered Backprojection type algorithms. It introduces non-horizontal filtering directions and a voxel specific backprojection operation in order to appropriately weight the acquired data. However, there is still missing frequency data due to the circular source trajectory. Even without employing additional algorithmic approaches to estimate the missing frequency data, significant improvements in the quantitative accuracy and the reproducibility of CT number values for different angular scan ranges has been achieved compared with the standard FDK based reconstruction. We note that the temporal artifact behavior and temporal resolution of the algorithm is the subject of future study.

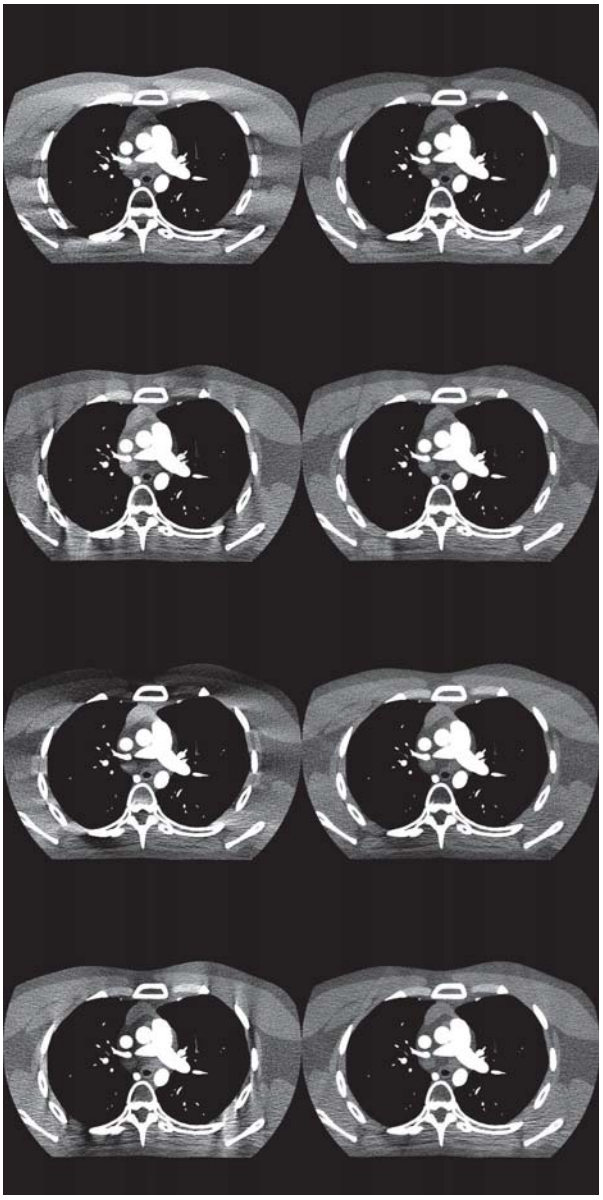


Fig. 5. Reconstructed images of a slice with cone angle $5.13°$. Left: P-FDK Reconstruction. Right: ACE Reconstruction. In each case the same range of view angles was input to the algorithms. The four different realization correspond to center view angles each separated by $90°$. [-200 200]HU

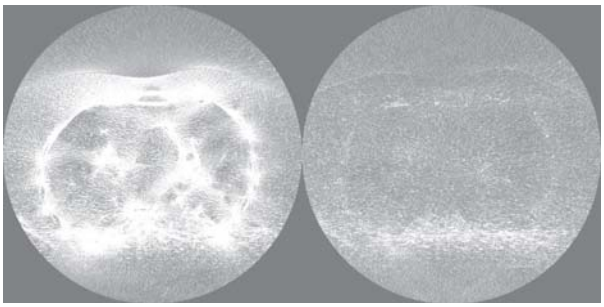


Fig. 6. Maps of the standard deviation in each pixel with respect to central view angle from the multiple acquisitions each separated by $15°$ (for slice with cone angle $5.13°$). Left: P-FDK Reconstruction. Right: ACE Reconstruction. [-50 50] HU

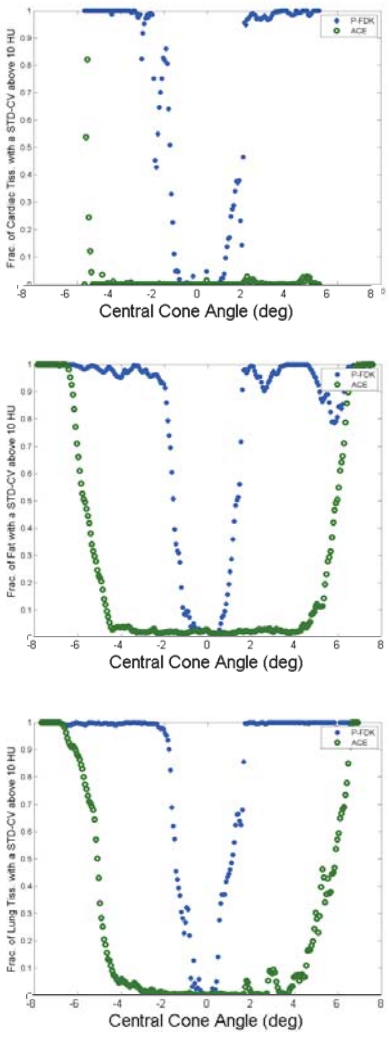


Fig. 7. For each of the automated ROIs from top to bottom: cardiac, fat and lung, the plots are given here of the fraction of voxels within that ROI which has a standard deviation with respect to view angle of less than 10 HU.

REFERENCES

[1] A. Katsevich, "A general scheme for constructing inversion algorithms for cone-beam CT," *Int. J. Math. Sci.*, vol. 21, pp. 1305–1321, 2003.

[2] G. H. Chen, "An alternative derivation of katsevich's cone-beam reconstruction formula," *Med. Phys.*, vol. 30, pp. 3217–3226, 2003.

[3] B. E. Nett, T. L. Zhuang, S. Leng, and G.-H. Chen, "Arc based cone-beam reconstruction algorithm using an equal weighting scheme," *Journal of X-Ray Science and Technology*, vol. 15, pp. 19–48, 2007.

[4] F. Noo, J. Pack, and D. Heuscher, "Exact helical reconstruction using native cone-beam geometries," *Phys. Med. Biol.*, vol. 48, pp. 3787–3818, 2003.

[5] D. L. Parker, "Optimal short-scan convolution reconstruction for fan beam CT," *Med. Phys.*, vol. 9, pp. 254–257, 1982.

Full Cone-Beam Consistency Conditions for Sources on a Plane

Rolf Clackdoyle and Laurent Desbat

***Abstract*-- Full (necessary and sufficient) consistency conditions are presented for cone-beam projections with sources on a plane. The object support is assumed to lie entirely on one side of the source plane. We have also established full consistency conditions for planograms and linograms, in both parallel and divergent beam formats. We show that moments of the appropriately weighted cone-beam projections form polynomials in the source variables, similar to the Helgason-Ludwig conditions. The degree of the polynomial matches the degree of the moment. All the consistency conditions stated here appear to be new. A simulation example is presented for the circular tomosynthesis geometry.**

I. Introduction

In image reconstruction from projections, consistency conditions (also known as range conditions) are mathematical expressions that describe the crosstalk of information between measured projections. Consistency conditions have been widely used in reconstruction algorithms for a range of medical imaging applications, with dozens of publications over the past 20 years in SPECT (e.g. [Nat93a] [Gli94] [Men99] [Erl00]), in PET (e.g. [Def95] [Wel03] [Lay05] [Def12]), and in X-ray CT (e.g. [Bas00] [Pat02] [Hsi04] [Yu07] [Tan11]). A typical approach is to use consistency to identify the parameters of some systematic effect in the imaging model such as rigid motion parameters, a photon attenuation coefficient, a beam hardening scaling factor, an elliptical body outline, amongst various possibilities. So for many applications, a collection of necessary conditions on the projections are needed in some convenient format for processing.

Consistency conditions for parallel projections in two dimensions are well known and take different forms such as the Helgason-Ludwig (HL) conditions [Lud66] [Hel80] or the frequency-distance relation on the Fourier transform of a sinogram [Edh86]. For applications in SPECT, conditions for the exponential ray transform (also called the exponential X-ray transform) are known [Agu95], and necessary conditions are also known for the two-dimensional (2D) attenuated Radon transform [Nat83].

For cone-beam or fanbeam projections, much less is known. Of the various publications on range conditions for divergent projections (e.g. [Fin83a] [Fin83b] [Pat02] [Che05] [Yu06] [Lev10] [Cla13]) most (not all) of them are just the parallel conditions re-expressed using fanbeam or cone-beam variables. This approach presents the disadvantage that a complete set of projections must be available so that the underlying parallel geometry is completely sampled. Some of the other formulations, not related to the parallel case also require a complete set of projections [Lou89] [Nat93b] [Maz10]. For applications, it is useful to have a method of processing a small finite set of projections, so the conditions should be amenable to this preference. A discussion of this point can be found in [Cla13].

In this work we are considering cone-beam projections, and we restrict our attention to planar source trajectories with the object support being entirely on one side of the trajectory plane. Currently, consistency conditions for this geometry can be obtained by applying John's condition (see [Fin85] [Pat02]) which is a partial differential equation with the disadvantage that it only treats local information in the projections; or by considering restrictions of the cone-beam geometry to fanbeam cases [Lev10]; or by using Grangeat's result [Gra91] which also only uses lines of data on the cone-beam projections. The conditions described below treat full cone-beam projections and have the added appeal of being in the familiar form of moments of the projections, similar to the HL conditions for parallel projections. We also indicate that the consistency conditions can easily be converted to planogram coordinates [Bra04] and we give full (necessary and sufficient) consistency conditions for planogram projections in both cone-beam and parallel formats. Similar results for linograms [Edh87] are readily extracted from the existing literature and will be stated below too.

II. Theory

A. Cone-beam consistency for a planar source trajectory

Let the source trajectory be $\underline{a}(\lambda)$ and without loss of generality, define the coordinate system so that the trajectory plane is $z = 0$ and the object lies in the $z > 0$ half-space. The x and y axes can be chosen freely in the $z = 0$ plane. The unit vector $\underline{\gamma}$ is selected using conventional ϕ, θ coordinates, so $\underline{\gamma}_{\phi,\theta} = (\cos\phi\sin\theta,\ \sin\phi\sin\theta, \cos\theta)$. The cone-beam projection $g(\underline{a}, \cdot)$ is given by

$$g(\underline{a}, \underline{\gamma}_{\phi,\theta}) = \int_0^\infty f(\underline{a} + r\underline{\gamma}_{\phi,\theta})dr \qquad (1)$$

For cone-beam consistency conditions, we first define

R. Clackdoyle is with the Laboratoire Hubert Curien, CNRS UMR 5516, Saint Etienne, France (e-mail: rolf.clackdoyle@univ-st-etienne.fr).

L. Desbat is with the TIMC-IMAG laboratory, CNRS UMR 5525, and Joseph Fourier University, Grenoble, France (e-mail laurent.desbat@imag.fr).

This work was partially supported by the Agence Nationale de la Recherche (France), project "DROITE," number ANR-12-BS01-0018.

$$J_n(\underline{a}, U, V) = \iint g(\underline{a}, \underline{\gamma}_{\phi,\theta})(U\cos\phi + V\sin\phi)^n \frac{\tan^{n+1}\theta}{\cos\theta} d\phi d\theta \quad (2)$$

for each $n = 0, 1, 2\ldots$. The cone-beam projections g will then satisfy

$$J_n(\underline{a}, U, V) = R_n(U, V, -a_x U - a_y V) \quad (3)$$

for all $\underline{a}$ in the plane, where for each n, $R_n(X, Y, Z)$ is a homogeneous polynomial of degree n.

These conditions are easily established. Simply substituting equation 1 into the expression for J_n given by equation 2, and changing from spherical to cartesian coordinates $\underline{a} + r\underline{\gamma}_{\phi,\theta} = (x, y, z)$, with $r^2 \sin\theta dr d\phi d\theta = dxdydz$, and recalling that $a_z = 0$, we quickly obtain

$$J_n(\underline{a}, U, V) = \iiint \frac{f(x, y, z)}{z^{n+2}}(xU + yV - a_x U - a_y V)^n dxdydz$$
$$= R_n(U, V, -a_x U - a_y V) \quad (4)$$

where $R_n(X, Y, Z) = \sum_{i+j+k=n} c_{i,j,k} X^i Y^j Z^k$, and

$$c_{i,j,k} = \frac{n!}{i!\, j!\, k!} \iiint f(x, y, z) \frac{x^i y^j}{z^{n+2}} dxdydz \quad (5)$$

so R_n is a homogeneous polynomial of degree n as claimed.

These consistency conditions are *full* because the converse is also true. Given some projection function g, which satisfies equation 3 for each non-negative integer n and each $\underline{a}$ in the plane, and where R_n is a homogeneous polynomial of degree n, there exists some function f such that equation 1 is satisfied. Our proof of this fact appeals to several existing theorems in image reconstruction theory and is too long to be presented in this abstract. To the best of our knowledge, these conditions are new.

For a flat detector parallel to the trajectory plane, these consistency conditions can be written in simpler form as will be shown below.

B. Full consistency conditions for planograms

Planogram coordinates [Bra04] are a three-dimensional (3D) version of linogram coordinates [Edh85]. As shown in fig. 1, planogram coordinates are suitable for a (conceptual) PET system consisting of two parallel infinite flat detectors. An integration line is specified by absolute coordinates on one detector and relative coordinates on the second detector. For convenience it is assumed that the distance between the detectors is one (which is not restrictive because simply scaling the units will accomplish this). We will assume that the first detector is in the $z = 0$ plane and the second detector in the $z = 1$ plane. With an object f of compact support situated between the detectors we define the planogram by

$$\hat{p}(x, y, u, v) = \int f((x, y, 0) + r(u, v, 1))dr \quad (6)$$

Equation 6 is written $\hat{p} = \hat{P}f$ for short. The reason for the "hat" is to emphasize that the planograms are in cone-beam format. See fig. 1(a). For each source location (x, y) on the first detector, the cone-beam projection $\hat{p}(x, y, \cdot, \cdot)$ is given by equation 6. A parallel projection version is defined below.

Equations 1 and 6 are linked by associating $\underline{a} = (x, y, 0)$, and $(\cos\phi\tan\theta, \sin\phi\tan\theta, 1) = (u, v, 1)$. It is then straightforward to verify that

$$\cos\theta\; g(\underline{a}, \underline{\gamma}_{\phi,\theta}) = \hat{p}(x, y, u, v) \quad (7)$$

and the cone-beam consistency conditions can be readily converted to full planogram consistency conditions which have a much simpler form:

P1: Let $\hat{Q}_n(x, y, U, V) = \iint \hat{p}(x, y, u, v)(uU + vV)^n du dv$. Then $\hat{p} = \hat{P}f$ for some f if and only if for all $n = 0, 1, 2\ldots$

$$\hat{Q}_n(x, y, U, V) = \hat{R}_n(U, V, -xU - yV) \quad (8)$$

where $\hat{R}_n$ is a homogenous polynomial of degree n.

Turning now to planograms in parallel format, the variables (u, v) are the projection indices, indicating the direction of the projection, and the individual ray variables are now (x, y) (see fig. 1(b)). The idea is that a parallel projection $\bar{\bar{p}}(u, v, \cdot, \cdot)$ with direction (u, v) is measured on the first detector. We write $\bar{\bar{p}} = \bar{\bar{P}}f$ for the equation

$$\bar{\bar{p}}(u, v, x, y) = \int f((x, y, 0) + r(u, v, 1))dr \quad (9)$$

so $\bar{\bar{p}}(u, v, x, y) = \hat{p}(x, y, u, v)$. However, it is important to recognize that these are fundamentally different when considered as functions of their projection indices (the first 2 variables). Now theorem 4.3 in [Nat86] provides full consistency for parallel projections in three-dimensions, and a simple change of variables provides the following result:

P2: Let $\bar{\bar{Q}}_n(u, v, X, Y) = \iint \bar{\bar{p}}(u, v, x, y)(xX + yY)^n dxdy$. Then $\bar{\bar{p}} = \bar{\bar{P}}f$ for some f if and only if for all $n = 0, 1, 2\ldots$

$$\bar{\bar{Q}}_n(u, v, X, Y) = \bar{\bar{R}}_n(X, Y, -uX - vY) \quad (10)$$

where $\bar{\bar{R}}_n$ is a homogenous polynomial of degree n.

The remarkable symmetry of the two results P1 and P2 belies their mathematical equivalence. We found that proving one from the other is not direct, and required the machinery of [Edh96]. Even for the same *f*, the polynomials $\hat{R}_n$ and $\bar{\bar{R}}_n$ are not the same in general.

C. Full consistency conditions for linograms

We present the corresponding linogram results, which are

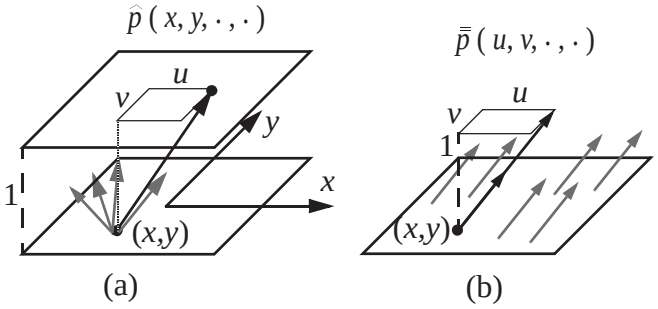


Fig 1. Planograms. (a) Cone-beam planograms. The source point (x, y) specifies the projection. The rays within a projection are specified by direction (u, v). (b) Parallel projection planograms. The direction of the projection is (u, v). The projection rays are specified by their intersection point with the detector (x, y).

the 2D versions of P1 and P2 above. Conceptually the two detector lines are vertical, separated by a distance of one, and the object lies between them.

For linograms in fanbeam format, we define $\hat{l} = \hat{L}f$ by

$$\hat{l}(y, u) = \int f((0, y) + r(1, u))dr \qquad (11)$$

and an elementary substitution of variables applied to theorem 1 in [Cla13] immediately gives us

L1: Let $\hat{K}_n(y) = \int \hat{l}(y, u)u^n du$. Then $\hat{l} = \hat{L}f$ for some f if and only if for each $n = 0, 1, 2\ldots$, the function $\hat{K}_n$ is a polynomial of degree n. (I.e. $\hat{K}_n(y) = c_0 + c_1 y + \ldots c_n y^n$.)

For linograms in parallel format, we define $l = Lf$ by

$$l(u, y) = \int f((0, y) + r(1, u))dr \qquad (12)$$

and we note the misleadingly simple $l(u, y) = \hat{l}(y, u)$. From the well-known 2D HL conditions, full consistency for parallel linograms are easily derived and are given by

L2: Let $K_n(u) = \int l(u, y)y^n dy$. then $l = Lf$ for some f if and only if for each $n = 0, 1, 2\ldots$, the function K_n is a polynomial of degree n. (I.e. $K_n(u) = c_0 + c_1 u + \ldots c_n u^n$.)

The symmetry in the definitions and the stated results suggest a much more trivial mathematical equivalence than actually exists between the fanbeam and parallel linograms.

III Simulations

A. Circular tomosynthesis geometry

Because recent advances have made progress with straight-line source trajectories in cone-beam tomography [Lev10] [Cla13], we chose to simulate a circular trajectory. This scanning configuration was probably first known as circular tomosynthesis. Figure 2 illustrates the geometry and an example projection of a high-contrast version of the 3D Shepp-Logan phantom. The projections were simulated using analytic line-length computations on the component ellipsoids. The detector was fixed (did not move with or conjugate to the source), and had 1024 x 1024 pixels. Thirty six projections were taken at 10^o increments along the source trajectory; this was enough projections to validate the theory and demonstrate the concepts. The cone-angle was a substantial 20^o (41^o full-angle) to ensure a strong cone-beam effect and avoid being perceived as 'nearly parallel.'

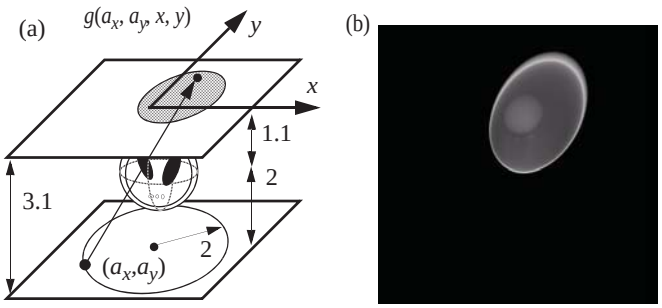


Fig 2. Circular tomosynthesis geometry. (a) the source (a_x, a_y) lies on a circle of radius 2 on the $z = 0$ plane. The center of the 3D Shepp Logan phantom is in the $z = 2$ plane, and the detector is in the $z = 3.1$ plane. Detector coordinates are (x, y). (b) One of the 36 simulated cone-beam projections.

B. Full consistency for fixed detector systems

The cone-beam measurements are given by $g(\underline{a}, \cdot)$ as specified in equation 1. We immediately scale and weight the projections so that the detector distance is one, and $\hat{g}(\underline{a}, x, y) = \cos\theta\, g(\underline{a}, \gamma_{\phi,\theta})$ with the variables linked as follows: $x - a_x = \cos\phi\tan\theta$ and $y - a_y = \sin\phi\tan\theta$. The "hat" on the g is a reminder both that the detector is fixed, and more importantly, that the raw projection measurements have been multiplied by $\cos\theta$, which is the cosine of the angle of incidence of the ray (θ is the angle between the detector normal and the incoming ray). This $\cos\theta$ term often appears in reconstruction algorithms for cone-beam tomography. To compress the notation slightly, we now use (a, b) instead of (a_x, a_y). It is easily shown that

$$\hat{g}(a, b, x, y) = \int f((a, b, 0) + r(x - a, y - b, 1))dr \qquad (13)$$

which is similar to the definition of $\hat{p}$ from equation 9.

Full consistency conditions for $\hat{g}(a, b, \cdot, \cdot)$ can be written in the same form as for the planograms case. Defining $\hat{J}_n(a, b, X, Y) = \iint \hat{g}(a, b, x, y)(xX + yY)^n dxdy$, it can be shown that $\hat{g}$ satisfies equation 13 for some f if and only if, for all n, the equality

$$\hat{J}_n(a, b, X, Y) = \hat{R}_n(X, Y, -aX - bY) \qquad (14)$$

holds for all (a, b) for some homogeneous polynomial $\hat{R}_n$ of degree n. (This function $\hat{R}_n$ is unrelated to the planogram $\hat{R}_n$)

C. Moment conditions on the projection data

To use the consistency conditions in practice, we first define the i-j th moment of the (cosine-scaled) projection by

$$M_{ij} = \iint \hat{g}(a, b, x, y)\, x^i\, y^j\, dxdy \qquad (15)$$

and we note that for any measured projection, the number M_{ij} is easy to compute (for any pair (i, j)). Note also from the definition of $\hat{J}_n$ that (dropping the (a, b))

$$\hat{J}_n(X, Y) = c_0 Y^n + c_1 XY^{n-1} + c_2 X^2 Y^{n-2} + \ldots c_n X^n$$

$$c_k = \binom{n}{k} M_{k, n-k} \qquad (16)$$

Also, the function $\hat{R}_n$ is of the form

$$\hat{R}_n(X, Y, Z) = \sum_{i+j+k=n} c_{i,j,k}\, X^i\, Y^j\, Z^k \qquad (17)$$

Now, setting $Z = -aX - bY$ and equating coefficients of $X^i Y^j$ in equations 16 and 17 leads to specific consistency conditions in terms of M_{ij}.

First, for $n = 0$, we immediately have $M_{00} = c_{000}$, so

$$\iint \hat{g}(a, b, x, y)\, dxdy = c_{000} \qquad (n = 0)$$

The sum of the (cosine-weighted) cone-beam projections is the same for all projections.

For $n = 1$, we obtain $M_{10} = c_{100} - c_{001}a$ and $M_{01} = c_{010} - c_{001}b$ so

$$\iint \hat{g}(a, b, x, y)\, x\, dxdy = -c_{001}a + c_{100}$$

$$\iint \hat{g}(a, b, x, y)\, y\, dxdy = -c_{001}b + c_{010} \qquad (n = 1)$$

For $n = 2$ a similar analysis yields

$$\iint \hat{g}(a,b,x,y)\, x^2 dxdy = c_{002}a^2 - c_{101}a + c_{200}$$

$$\iint \hat{g}(a,b,x,y)\, y^2 dxdy = c_{002}b^2 - c_{011}b + c_{020} \qquad (n = 2)$$

$$\iint \hat{g}(a,b,x,y)\, xy\, dxdy = c_{002}ab - \frac{c_{011}}{2}a - \frac{c_{101}}{2}b + \frac{c_{110}}{2}$$

In summary

- M_{ij} is a polynomial of degree $i+j$ in (a,b)
- M_{n0} is a polynomial of degree n in a (resp. M_{0n} in b)
- the coefficients for each $\hat{R}_n$ can be found by fitting polynomials of degree n in (a,b) to each M_{ij} with $i+j=n$.
- for each n, the coefficients of the polynomials in (a,b) intertwine

D. Simulation results

We illustrate in fig. 3 the polynomials fit from the 36 simulated projections for the cases M_{00}, M_{10}, M_{01}, M_{20}, M_{02}. As expected for noise-free data, the fits are virtually perfect (accurate to under 0.1%), and verify the consistency theory.

V. Discussion and Conclusions

We have derived new cone-beam consistency conditions for sources lying on a plane. For consistent data the moments of the projections must be polynomials in the source variables, of the appropriate degree. Our simulations with ideal data were in agreement with the theory.

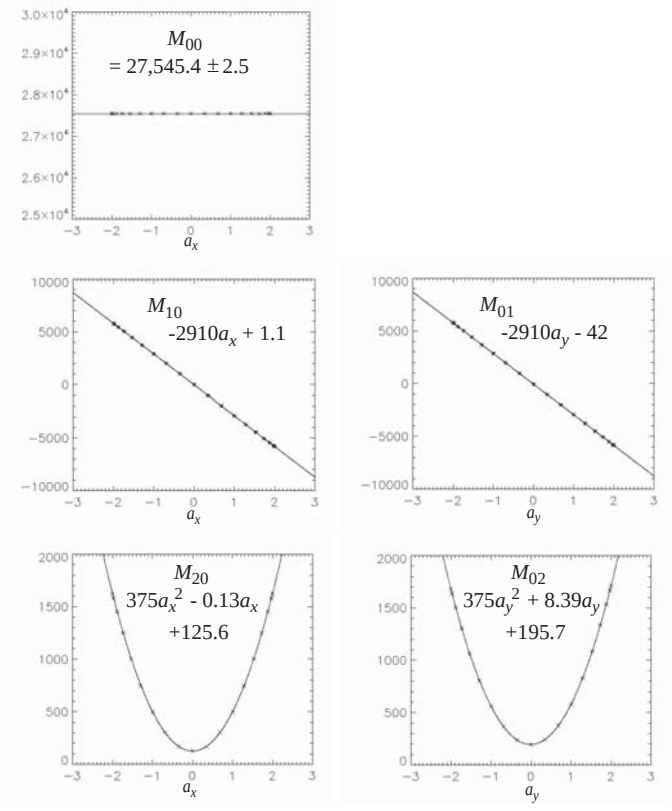


Fig 3. Projection moments, with polynomial fits. The horizontal axis is the source position. The 36 calculated values for each moment are plotted with stars (∗) and the least-squares fitted polynomial is drawn with a solid line. The results match the theory. Note that M_{10} and M_{01} have the same slope, as predicted.

The presented consistency conditions are necessary and sufficient so in principle, other conditions can be derived from them. However we have not yet linked our work mathematically to the conditions of John or Grangeat.

VI. References

[Agu95] V. Aguilar and P. Kuchment. "Range conditions for the multi-dimensional exponential x-ray transform." *Inv. Probs* 11, 977-982, 1995.

[Bas00] S. Basu and Y. Bresler. "Uniqueness of tomography with unknown view angles." *IEEE Trans Imag Proc* 9, 1092-1106, 2000.

[Bra04] D. Brasse, P.E. Kinahan, R. Clackdoyle, M. Defrise, C. Comtat and D.W. Townsend. "Fast Fully 3-D Image Reconstruction in PET Using Planograms" *IEEE Trans Med Imag* 23, 413-425, 2004.

[Cla13] R. Clackdoyle. "Necessary and Sufficient Consistency Conditions for Fanbeam Projections along a Line." *IEEE Trans Nucl Sci* (in press) 2013.

[Che05] G.-H. Chen and S. Leng. "A new data consistency condition for fan-beam projection data." *Med Phys* 32, 961-967, 2005.

[Def95] M. Defrise. "A factorization method for the 3D x-ray transform." *Inv Probs* 11, 983-994, 1995.

[Def12] M. Defrise, A. Rezaei and J. Nuyts. "Time-of-flight PET data determine the attenuation sinogram up to a constant." *Phys. Med. Biol*. 57, 885-899, 2012.

[Edh86] P.R. Edholm, R.M. Lewitt and B. Lindholm. "Novel properties of the Fourier decomposition of the sinogram." *Proc SPIE* 671, 8-18, 2006.

[Edh87] P.R. Edholm and G.T. Herman. "Linograms in image reconstruction from projections." *IEEE Trans. Med. Imag*. MI-6, 301–307, 1987.

[Edh96] P.R. Edholm and P.-E. Danielsson. "A theorem on divergent projections." *Three Dimensional Image Reconstruction in Radiation and Nuclear Medicine*. P Grangeat and J L Amans eds (Amsterdam: Kluwer) 35-45, 1996.

[Erl00] K. Erlandsson, D. Visvikis, W.A. Waddington and P. Jarritt. "Truncation Reduction in Fan-Beam Transmission Scanning Using the Radon Transform Consistency Conditions." *IEEE Trans. Nuc. Sci*. 47, 989–993, 2000.

[Fin83a] D.V. Finch and D.C. Solmon. "A characterization of the range of the divergent beam X-ray transform." *SIAM J Mat. Anal* 14, 767–771, 1983.

[Fin83b] D.V. Finch and D.C. Solmon. "Sums of homogeneous function and the range of the divergent beam X-ray transform." *Numer Func Anal and Optimiz* 5, 363–419, 1983.

[Fin85] D.V. Finch. "Cone beam reconstruction with sources on a curve." *SIAM J. Appl. Math*. 45, 665–673, 1985.

[Gli94] S.J. Glick, B.C. Penney, M.A. King and C.L. Byrne "Noniterative Compensation for the Distance-Dependent Detector Response and Photon Attenuation in SPECT Imaging" *IEEE Trans.Med. Imag*. 13, 363–374, 1994.

[Gra91] P. Grangeat. "Mathematical framework of cone beam 3D reconstruction via the first derivative of the Radon transform." *Mathematical Methods in Tomography (Lecture Notes in Mathematics, vol 1497)* ed Herman, Louis and Natterer (New York: Springer), 66–97, 1991.

[Hel80] S. Helgason. *The Radon Transform*. (Boston: Birkhauser), 1980.

[Hsi04] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash and T.J. Myers. "A novel reconstruction algorithm to extend the CT scan field-of-view." Med. Phys. 31, 2385–2391, 2004.

[Lay05] C.M. Laymon, M. Swadley, J.E. Bowsher, T. Blodgett and S.K. Ziolko. "Evaluation of Sinogram Consistency Conditions for Identifying Artifactual PET Attenuation Images". *Conf Rec of the 2005 Nucl Sci Symp and Med Imag Conf*. 3, 1611-1614, 2005.

[Lev10] M.S. Levine, E.Y. Sidky and X. Pan. "Consistency Conditions for Cone-Beam CT Data Acquired with a Straight-Line Source Trajectory." *Tsinghua Sci Technol* 15,56–61, 2010.

[Lou89] A K. Louis and A. Rieder. "Incomplete Data Problems in X-Ray Computerized Tomography II. Truncated Projections and Region-of-Interest Tomography." *Numer. Math.* 56, 371–383, 1989.

[Lud66] D. Ludwig. "The Radon transform on Euclidean space." *Comm. Pure Appl. Math*. 19, 49–81, 1966.

[Maz10] S.R. Mazin and N.J. Pelc. "Fourier properties of the fanbeam sinogram." *Med. Phys* 37, 1674-1680, 2010.

[Men99] C. Mennessier, F. Noo, R. Clackdoyle, G. Bal and L. Desbat. "Attenuation correction in SPECT using consistency conditions for the exponential ray transform." *Phys. Med. Biol*. 44, 2483–2510, 1999.

[Nat83] F. Natterer. "Computerized tomography with unknown sources." *SIAM J. Appl. Math*. 43, 1201-1212, 1983.

[Nat86] F. Natterer. *The Mathematics of Computerized Tomography.* (New York: Wiley-Teubner), 1986.

[Nat93a] F. Natterer. "Determination of tissue attenuation in emission tomography of optically dense media." *Inv. Probs* 9, 731–736, 1993.

[Nat93b] F. Natterer. "Sampling in fan beam tomography." *SIAM J Appl Math* 53, 358-380, 1993.

[Pat02] S.K. Patch. "Consistency conditions upon 3D CT data and the wave equation." *Phys. Med. Biol*. 47, 2637-2650, 2002.

[Tan11] S. Tang, X. Mou, Q. Xu, Y. Zhang, J. Bennett and H. Yu. "Data consistency condition-based beam-hardening correction." *Opt. Eng.* 50, 076501(1–13), 2011.

[Wel03] A. Welch, W. Hallett, P. Marsden and A. Bromiley "Accurate Attenuation Correction in PET using Short Transmission Scans and Consistency Information." *IEEE Trans Nucl Sci* 50, 427–432, 2003.

[Yu06] H. Yu, Y. Wei, J. Hsieh and G. Wang. "Data Consistency Based Translational Motion Artifact Reduction in Fan-Beam CT." *IEEE Trans. Med. Imag*.25, 792–803, 2006.

[Yu07] H. Yu and G. Wang. "Data Consistency Based Rigid Motion Artifact Reduction in Fan-Beam CT." *IEEE Trans. Med. Imag*. 26, 249–260, 2007.

Task-Based Trajectories in Iteratively Reconstructed Interventional Cone-Beam CT

J. Webster Stayman and Jeffrey H. Siewerdsen

***Abstract*—Interventional imaging scenarios are rich in prior knowledge of patient anatomy (e.g., preoperative CT and/or successive CBCT scans) and tend to have a specific and well-defined imaging tasks. This presents an opportunity to integrate such information into the image acquisition process by means of a customized CBCT scan orbit in which the source-detector trajectory is chosen based on a knowledge of the task and anatomical context in a manner to maximize performance. We adopt task-based performance predictors based on a numerical (non-prewhitening matched filter) observer model and approximations to the local noise and spatial resolution properties of penalized-likelihood reconstruction. These predictions are then used to identify projections that maximize task performance, beginning with the projection view that maximizes detectability, proceeding to the next-best view, and continuing in an (arbitrarily constrained) orbit that can be physically realized on advanced robotic C-arm platforms. We illustrate the approach in simulations considering a robotic C-arm capable of rotational orbits with oblique angulations to compute orbits that are optimal to a specified imaging task. We demonstrate the performance of task-based trajectory versus simple and complex orbits to illustrate the advantages of integrating prior knowledge and the imaging task into customized acquisitions.**

***Index Terms*—CT Reconstruction, Task-Based Detectability, Interventional Imaging, Cone-beam CT, Model-based Reconstruction, Projection View Optimization.**

I. Introduction

Advanced model-based reconstruction has shown a great deal of promise in its ability to improve noise-resolution tradeoffs over traditional approaches [1] and to accommodate low-fidelity data through exposure reductions or sparse projection sampling[2]. While such methods are now being adopted clinically for diagnostic CT, their use in interventional imaging (e.g., intraoperative CBCT) has yet to be fully explored. There are many aspects of interventional imaging that make it distinct from diagnostic imaging, including: 1) Virtually all patients undergoing image-guided interventions have pre-operative imaging studies offering a wealth of patient-specific anatomical prior knowledge; 2) Interventional systems are well-suited to sparse acquisitions and more general source-detector trajectories (e.g., non-circular orbits); and 3) Imaging tasks associated with interventional imaging are well-defined and highly focused for specific procedures (e.g., detection and localization of known targets). Thus, in many respects, interventional imaging is an ideal target for leveraging advanced model-based reconstruction approaches that accommodate arbitrary orbits and sparsity to customize data acquisitions based on prior knowledge of the patient for maximal performance in specific imaging tasks.

This work was supported in part by NIH Grants No. 2R01-CA-112163 and R21-EB-014964.

J. Webster Stayman is with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21212 USA (phone: 410-955-1314; fax: 410-955-1115; e-mail: web.stayman@jhu.edu).

Recent attempts to define acquisitions for increased performance with iterative reconstruction include selection of the number of projections based on condition number [3] and methods that choose projections based on edge content [4]. In this paper, we introduce a framework that leverages previously acquired patient-specific anatomical information to predict the task-based performance of a model-based penalized-likelihood reconstruction for given source-detector trajectories. We then use this framework to select the most information-rich projections that maximize task performance predictions and then assemble a customized task- and patient-specific source-detector trajectory subject to physical / geometrical constraints of the source-detector (robotic C-arm) platform.

II. Methods

A. System Models

Both iterative reconstruction and performance predictors require a system model that must be general enough to accommodate particular acquisition geometries and include measurement noise to leverage the advantages of statistical approaches. We adopt the following general vectorized forward model for the mean measurements:

$$\bar{y} = \mathbf{D}\{b\}\exp\left(-\mathbf{A}\mu\right), \qquad (1)$$

where the measurement vector, y, is related to the volume μ through Beer's Law and includes measurement-dependent gains in the diagonal matrix $\mathbf{D}\{b\}$. Each element of the system matrix, $\mathbf{A}$, models the contribution of a specific voxel to a specific projection measurement. This model can accommodate arbitrary geometries like those obtainable with robotic C-arms (see Figure 1). We note that the system matrix for an entire orbit is comprised of smaller matrices for each 2D projection:

$$\mathbf{A}\begin{pmatrix}\{\theta_1,\phi_1\},\cdots,\\ \{\theta_N,\phi_N\}\end{pmatrix} = \begin{bmatrix}\mathbf{A}_{\theta_1,\phi_1}^T & \mathbf{A}_{\theta_2,\phi_2}^T & \cdots & \mathbf{A}_{\theta_N,\phi_N}^T\end{bmatrix}^T. \qquad (2)$$

In the current work, we focus on a system orbit parameterized by two angles, rotation angle (θ) and obliquity angle (ϕ).While the two angle scenario is an interesting subset

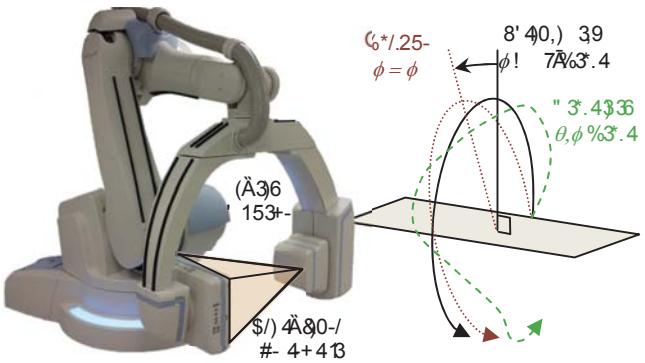


Figure 1: Robotically controlled C-arms are capable of a wide range of source-detector positions. Traditional cone-beam CT acquisitions are typically acquired over a standard 180°+fan angle orbit with no inclination; however, robotic C-arms allow oblique orbits or arbitrarily complex variations in rotation angle (θ) and obliquity (ϕ) throughout the orbit.

of achievable projections, robotic C-arms are capable of translations and modifications of the source-detector distance yielding many more possibilities.

B. Penalized-Likelihood Reconstruction

Once a specific forward model and trajectory has been chosen, it is straightforward to adopt a penalized-likelihood estimation approach for reconstruction. Unlike many analytic approaches, arbitrary trajectories and data sparsity are handled inherently, without modification of weighting factors, etc., once **A** has been defined. We consider the following estimator:

$$\hat{\mu} = \arg\max_{\mu} L(\mu; y) - \beta R(\mu), \tag{3}$$

that adopts a Poisson log-likelihood, L, and a quadratic penalty, $R(\mu)=\mu^T \mathbf{R}\mu$. A separable paraboloidal surrogates approach [5] is applied to iteratively solve (3).

C. Performance Prediction

The estimator in (3) is convenient, since one may write approximate predictors for the local point spread function (PSF) and local covariance [6], or equivalently, the local modulation transfer function (MTF) and local noise-power spectrum (NPS) [7]. Such imaging performance metrics are prevalent in image quality assessment and are leveraged here directly toward the 3D image acquisition and reconstruction process. From [6], the local PSF and covariance are approximately:

$$\begin{aligned} PSF_j &\approx \left[\mathbf{A}^T\mathbf{D}\mathbf{A} + \beta\mathbf{R}\right]^{-1}\mathbf{A}^T\mathbf{D}\mathbf{A}e_j \\ Cov\left\{\hat{\mu}_j\right\} &\approx \left[\mathbf{A}^T\mathbf{D}\mathbf{A} + \beta\mathbf{R}\right]^{-1}\mathbf{A}^T\mathbf{D}\mathbf{A}\left[\mathbf{A}^T\mathbf{D}\mathbf{A} + \beta\mathbf{R}\right]^{-1}e_j \end{aligned} \tag{4}$$

where $\mathbf{D}=\mathbf{D}\{y\}$ and e_j denotes a vector with unity j^{th} element and zero otherwise (specifying the location of interest as with a Kronecker delta function). Note that object-dependence enters (4) through **D,** which is dependent on the measurements. While (4) can be computed precisely using iterative approaches, such methods have high computational burden. Alternately, one may use a Fourier approximation to (4) as in [8]:

$$\begin{aligned} MTF_j &\approx \frac{\mathcal{F}\left\{\mathbf{A}^T\mathbf{D}\mathbf{A}e_j\right\}}{\mathcal{F}\left\{\mathbf{A}^T\mathbf{D}\mathbf{A}e_j + \beta\mathbf{R}e_j\right\}} \\ NPS_j &\approx \frac{\mathcal{F}\left\{\mathbf{A}^T\mathbf{D}\mathbf{A}e_j\right\}}{\left|\mathcal{F}\left\{\mathbf{A}^T\mathbf{D}\mathbf{A}e_j + \beta\mathbf{R}e_j\right\}\right|^2} \end{aligned}, \tag{5}$$

where $\mathcal{F}$ denotes a discrete Fourier transform and the divisions are element-by-element. With expressions for local MTF and NPS, one may predict estimator performance using a model observer. For example, using a non-prewhitening matched-filter observer one can express the detectability index [9, 10] as

$$d'^2 = \frac{\left[\iiint \left(MTF_j \cdot W_{Task}\right)^2 df_x df_y df_z\right]^2}{\iiint NPS_j \cdot \left(MTF_j \cdot W_{Task}\right)^2 df_x df_y df_z}, \tag{6}$$

where W_{Task} is the so-called task function given by the Fourier transform of the difference of two hypotheses (e.g., signal absent vs. signal present). Many other choices of numerical observer are possible, including those that more closely model the human visual system; however, the current initial investigations employ this simple model, which has demonstrated reasonable agreement with human observers in tomographic imaging relative to simple imaging tasks. [10] Thus, using (6), one may then predict performance for a given task (W_{Task}), object (via **D** in (5)), location (subscript j), and acquisition trajectory (**A** in (5)).

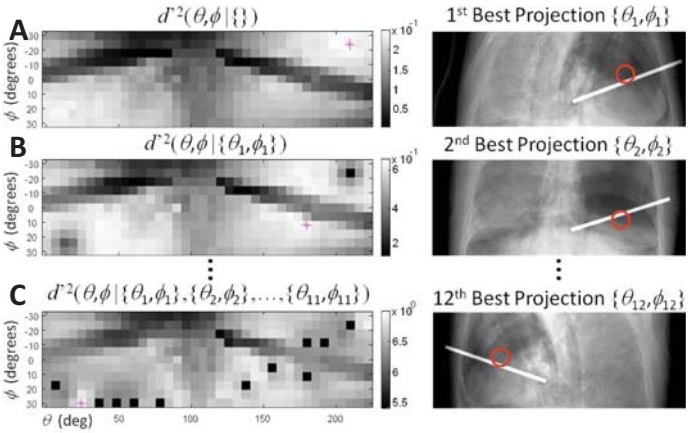


Figure 2: Illustration of the optimization approach used to find angle pairs that contribute to high detectability. The left column shows a detectability map for possible (rotation, obliquity) angle pair additions, the magenta asterisk identifies the new angle pair with highest detectability, and the right column illustrates the projection associated with that particular angle choice. (Since the low-contrast nodule is not visible in the projections, its location is identified with a red circle.) Three passes of the iterative design are shown including: A) Detectability given no previous angles. B) Detectability after addition of the first angle pair. C) Detectability after 11 angle pairs have been added. Note that "best" projections tend to be oblique angles that avoid nodule overlap with the dense surgical tool, and AP/PA projections (with decreased net attenuation) are preferred over lateral projections. Similarly, once an angle is chosen, subsequent detectability maps are decreased in the neighborhood of that angle (and the angle associated with the opposing view – see lower left of 2nd detectability map in (B)).

D. Methodology for Projection Selection

Because we are interested in finding the best source-detector trajectory for a given task and patient in an interventional setting where anatomical information is available from preoperative CT (or prior CBCT), we choose to optimize (6) over **A**. Recalling (2), we may substitute $\mathbf{A}=\mathbf{A}(\{\theta_1,\phi_1\},\ldots,\{\theta_N,\phi_N\})$ for different sets of projections into (6) to obtain $d'^2(\{\theta_1,\phi_1\},\ldots,\{\theta_N,\phi_N\})$.

The general optimization task is difficult, and performing a search over all possible combinations of N angles is prohibitive for larger N. Therefore, we introduce the following notation:

$$d'^2\left(\theta,\phi\middle|\{\theta_1,\phi_1\},\ldots,\{\theta_N,\phi_N\}\right), \tag{7}$$

which denotes a 2D function over θ and ϕ that expresses the overall detectability that a given projection angle yields when added to an orbit already containing a specified set of N projections. In other words, (7) yields a function whose maximum identifies the "next best projection view" based on task detectability. Thus, we may find highly performing sets of projections via a greedy approach where new angles are added to a growing set of projection angles starting with an empty set. That is, a set of projection angles is then constructed by iteratively finding the next most valuable projection $(\theta_{N+1},\phi_{N+1})$ in the detectability map and adding it to the existing set of N angle pairs. Stopping criteria may be formed based on number of angles, dose allocation, acquisition time, detectability, etc.

This greedy optimization approach is illustrated in Figure 2 for a simulated thoracic interventional imaging scenario. In this case, the patient anatomy contains a high-density surgical tool that is part of the intervention as well as a low-contrast pulmonary spherical nodule in a collapsed lung that is difficult to identify in projection images. The task function is the Fourier transform of the spherical nodule, and the location j is

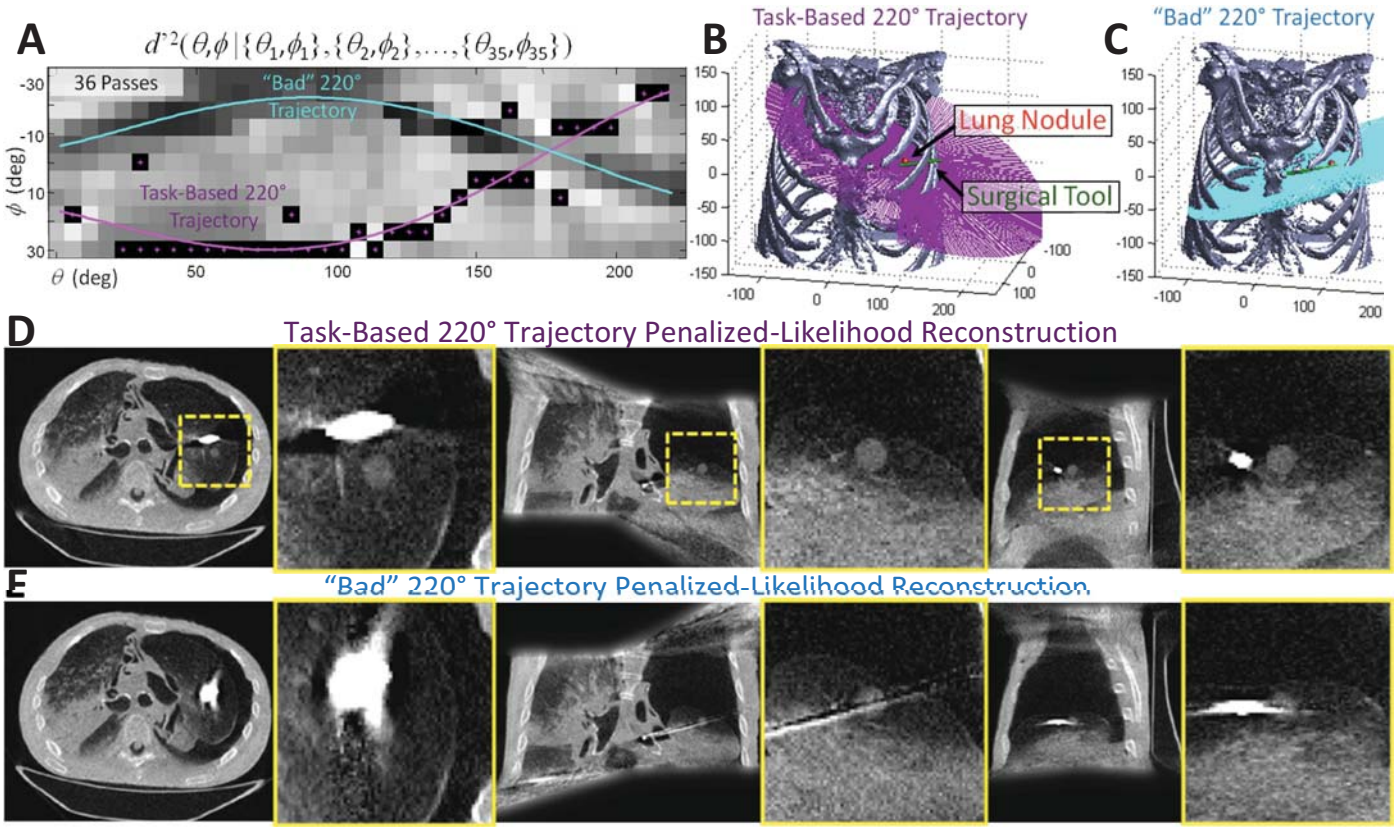


Figure 3: Illustration of a task-based source-detector trajectory using the proposed performance prediction approach. A) The detectability map after 36 passes of optimization. The "best" 36 angular positions are indicated with small magenta asterisks. To obtain a continuous orbit, a polynomial fit was made through these points (magenta curve). Another line fit through the detectability minima represents a particularly "bad" orbit (cyan curve). Representations of the task-based orbit (B) and "bad" orbit scenario (C) are also displayed, showing the source-detector ray that pierces the center of the lung nodule across all projections. Penalized-likelihood reconstruction of acquisitions based on the (D) task-based orbit and (E) "bad" orbit illustrate the relative performance of the two trajectories with all other acquisition and reconstruction parameters held constant. Note that the nodule is more easily detected in the task-based orbit, whereas severe blur around the high-density surgical tool and noise associated with poor data fidelity significantly degrade the image for the "bad" orbit.

matched to the true location of the nodule. Detectability is computed over a limited -30° to 30° obliquity and a 220° rotational range (samples every 6° in each direction); however, these limits may be adjusted to accommodate the mechanical capabilities and constraints for particular devices and interventional scenarios.

III. Results

We continue with the thoracic intervention scenario discussed in the previous section to illustrate the results of the proposed task-based trajectory design approach. In this investigation, a system with 1200 mm source-detector distance, 600 mm source-axis distance, and a 300x800 detector with 0.776 mm pixel pitch was simulated. A 300^3 volume (1 mm voxels) was used for all experiments, and an exposure equivalent of 10^3 photons per detector element was used for Poisson noise generation.

For these studies, we assumed that the image volume is known and perfectly registered for orbital trajectory design. To design a custom orbit for the spherical lung nodule detection task, we conducted 36 passes of the optimization approach detailed in the previous section. The final detectability map and selected angle pairs are shown in Figure 3A. Note that the selected projection angles tend to avoid overlap of the high-density surgical tool and the low-contrast nodule in projections. The influence of the surgical tool is apparent in the upper dark arc in the detectability map. Similarly, while lateral views are discouraged early in the optimization (dark vertical band in Figure 2), these views become important in later iterations (in Figure 3A this dark band is absent).

Interestingly, despite the lack of a constraint for a continuous trajectory, the task-based design has produced largely contiguous projection angle pairs. To produce a completely continuous orbit, we performed a polynomial curve fit through the 36 selected positions to produce a 220° orbit with 1° steps. A second "bad" orbit was also designed via a curve fit through regions of minimum detectability for comparison. These two orbits are illustrated in Figure 3B and C relative to the bony anatomy, lung nodule, and surgical tool. Note that while the task-based orbit avoids projections of the lung nodule that overlap with the surgical tool and bone, the "bad" trajectory is that in which the nodule and tool overlap in nearly every view.

The penalized-likelihood reconstructions associated with each trajectory are shown in Figure 3D and E. The low-contrast pulmonary nodule has qualitatively better detectability in the designed orbit, while the "bad" trajectory results in significantly increased blur and noise due to the low-fidelity data associated with projections through the high-density surgical tool. These effects are most pronounced in the axial images; however, significant streaking and noise is also apparent in the "bad" orbit in the sagittal and coronal slices.

A second experiment using the task-based trajectory design was conducted for a second task function. All simulation settings except for exposure (here set to 10^5 photons per detector element) remain the same as the previous experiment. In this scenario, we selected a task function based on a binary hypothesis test consisting of the same spherical nodule shifted

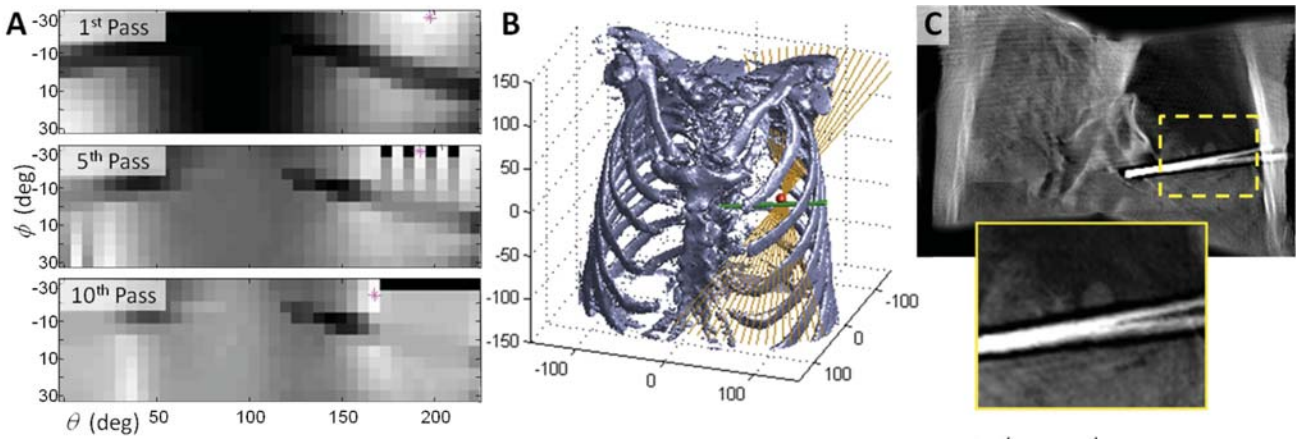


Figure 4: An application of the task-based trajectory design using a lateral localization task function. A) The 1st, 5th, and 10th passes of the optimization procedure show clustering of "next best" angles in the PA/AP direction with a high obliquity that avoids overlap of the nodule and surgical tool. B) A trajectory based on a 3° sampling interval on the best 10 angles is shown. C) Coronal views (appropriate for lateral localization) of a penalized-likelihood reconstruction show strong discrimination of the position of the nodule despite the limited-angle acquisition and small number of projection angles.

1 mm laterally in the patient, emulating a lateral localization task that is very different from the nodule detection task in the first experiment and emphasizing higher spatial frequencies in maximizing the detectability index.

The results of the lateral localization task are summarized in Figure 4. Figure 4A shows the detectability maps for the 1st, 5th, and 10th passes of the trajectory design. The maps reveal a number of interesting features. Not surprisingly, lateral projections contribute little to task performance for the lateral localization task (particularly in early iterations), and AP/PA views with high obliquities that separate the surgical tool and the low-contrast nodule are preferred. In addition, the initial angular pairs are spaced at intervals before filling in later iterations. A trajectory using a 3° sampling interval was fit to the best 10 angle pairs and is illustrated in Figure 4B. The resulting task-based trajectory for lateral localization is a limited angle acquisition (i.e., tomosynthesis). Coronal slices of a penalized-likelihood reconstruction are shown in Figure 4C. Even though the acquisition is highly limited in angle with a small number of views, the nodule is easily localized laterally based on this limited-angle reconstruction.

IV. Discussion

We have introduced a methodology for designing task-based trajectories for interventional imaging using penalized-likelihood reconstruction. The preliminary results show the potential for increased performance over standard orbits, including conventional circular orbits and (unlucky) oblique orbits that align patient anatomy, interventional tools, and/or the structure of interest in a manner that reduces detectability. There are a number of possible extensions to the proposed methodology including generalizations to more arbitrary geometries (e.g., including translation, magnification, etc.), optimization of detectability over a volume-of-interest instead of a single point, optimization of multiple task functions, limited field-of-view reconstructions, and addition of additional constraints (e.g., radiation dose and acquisition time). Another important challenge is an extension of these concepts to nonquadratic penalty approaches that are commonly used in CT.

One important practical consideration for this work is that the methodology presumes having a registered preoperative CT for design. Thus, a functional workflow must include a registration step, perhaps based on an initial projection image and 2D-3D registration. Since this registration is likely imperfect, there is also the possibility of an adaptive design technique that adjusts the orbit "on-the-fly." That is, as more information is obtained about the patient volume, the imaging system adapts its trajectory to maximize performance. The underlying framework presented in this paper is general and can accommodate much of the future work discussed above.

The proposed method leverages the wealth of information available in interventional imaging and combines it with methods of task-based performance evaluation to define optimal trajectories. Whereas traditional approaches tend to neglect the wealth of prior knowledge or use it in only a very coarse manner, the proposed framework integrates it fundamentally into the acquisition process. This is an important step in making imaging systems more aware of the objects they are imaging and the imaging tasks for which they intended, leading to increased imaging performance and potential reduction in dose.

References

[1]J. B. Thibault, *et al.*, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med Phys,* vol. 34, pp. 4526-44, Nov 2007.

[2]J. Bian, *et al.*, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Phys Med Biol,* vol. 55, pp. 6575-99, Nov 21 2010.

[3]E. Y. Sidky, *et al.*, "Sampling conditions for gradient-magnitude sparsity based image reconstruction algorithms," *Medical Imaging 2012: Physics of Medical Imaging,* vol. 8313, 2012.

[4]Z. Zheng and K. Mueller, "Identifying sets of favorable projections for few-view low-dose cone-beam CT scanning," in *11th Int't Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Potsdam, Germany, 2011, pp. 314-317.

[5]H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys Med Biol,* vol. 44, pp. 2835-51, Nov 1999.

[6]J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Trans Image Process,* vol. 5, pp. 1346-58, 1996.

[7]G. Gang, *et al.*, "Modeling and control of nonstationary noise characteristics in filtered-backprojection and penalized-likelihood image recostruction," in *SPIE Medical Imaging*, Orlando, 2013.

[8]J. W. Stayman and J. A. Fessler, "Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in fully 3-D SPECT," *IEEE Trans Med Imaging,* vol. 23, pp. 1543-56, Dec 2004.

[9]R. F. Wagner, *et al.*, "Application of information theory to the assessment of computed tomography," *Medical Physics,* vol. 6, pp. 83-94, 1979.

[10] G. J. Gang, *et al.*, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance," *Med Phys,* vol. 38, pp. 1754-68, Apr 2011.

Attenuation Correction Using a Single-View Mammogram for a 3D PET Scanner

PE Kinahan, C Zeng, LA Pierce II, KM Kanal, LR MacDonald
Department of Radiology, University of Washington, Seattle, WA USA

***Abstract*—We present an algorithm for attenuation correction using a single view mammogram for a 3D PET scanner. The need for an x-ray based attenuation correction (XAC) algorithm is motivated by the development of a small application-specific breast-imaging PET scanner (PET/X) that attaches to a full-field digital mammography system. The intended applications of the PET/X scanner require accurate quantitation, so attenuation correction is required.**

Methods: The XAC method is based on the assumption of constant attenuation at 511 keV, thus only an accurate estimate of the boundary of the breast under mild compression is needed. Most of the breast is bounded by the upper compression paddle and the lower support, which have a known separation, leaving only the edges with unknown thickness. By using the unprocessed mammogram, the attenuation length in these regions can be directly measured. This XAC method is tested with simulation studies to determine the errors in the estimated attenuation boundary as well as the impact on the accuracy of the reconstructed PET images. A simplistic breast shape composed of adipose tissue imaged at 30 keV was converted to 511 keV attenuation image. The corresponding PET emission data was corrected with perfect and estimated attenuation factors and lesion uptake was compared.

Results: The XAC method did not completely estimate the true attenuation boundary. This led to emission errors of approximately ±6%.

Conclusions: The XAC method is a potentially viable approach for attenuation correction with the PET/X scanner. Challenges may arise if non-biological materials are in the field of view. However additional information including un-attenuated PET images, dual-kVp or tomosynthesis data can potentially be used to reduce these errors. Further investigation is needed.

I. INTRODUCTION

Approximately 200,000 women in the U.S. are diagnosed with breast cancer annually [1]. Nearly all will receive some form of systemic therapy. As of 2007, there were 30 approved breast cancer therapies, the most of any cancer. Despite several biomarkers (e.g., tumor phenotype, receptor status) that are used to characterize the cancer and help determine treatment, efficacy is highly variable. More importantly, the success of therapy can only be determined after the patient has demonstrated significant changes in the size of their tumor, which usually does not occur until late in the course of treatment, i.e. after several months, and costs of tens of thousands of dollars. Early evaluation of response to therapy can be used to determine the efficacy of neoadjuvant therapy, i.e. therapy given in the week or two between initial diagnosis and removal of the primary tumor (Figure 1).

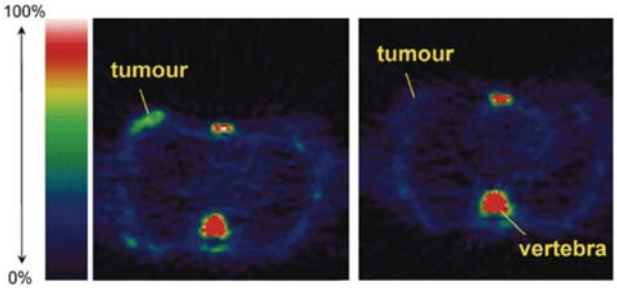


Figure 1. Quantitative FLT-PET images in a responding patient. (a) Pre-treatment and (b) 1 week into chemotherapy treatment images of a patient with grade II lobular carcinoma that responded to treatment [2].

This early assessment can provide valuable early information on the efficacy of adjuvant therapy that is planned to be used after surgery, which is the case for most breast cancer patients.

We are developing a compact positron emission tomography (PET) imaging module that can be used as an add-on with full-field X-ray mammography scanners (Figure 2) [3]. The PET/X scanner is designed for use with breast cancer patients to select and monitor the choice of the most effective therapy. Two essential components needed for using the PET/X scanner in accurate serial studies are anatomical localization (so the same location is imaged and/or biopsied) and attenuation correction (to accurately measure FDG uptake in the lesion). We present in the next section an algorithm for attenuation correction using a single-view mammogram.

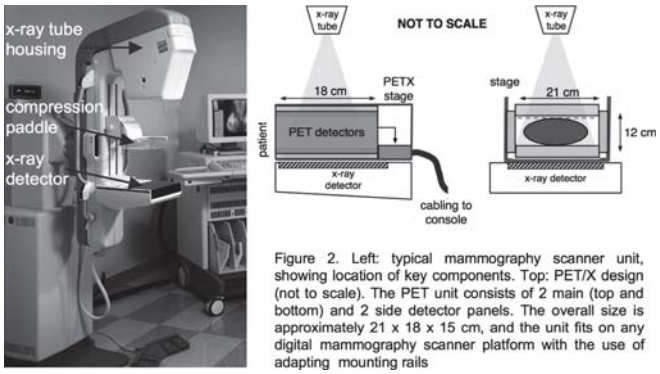


Figure 2. Left: typical mammography scanner unit, showing location of key components. Top: PET/X design (not to scale). The PET unit consists of 2 main (top and bottom) and 2 side detector panels. The overall size is approximately 21 x 18 x 15 cm, and the unit fits on any digital mammography scanner platform with the use of adapting mounting rails

II. THE X-RAY BASED ATTENUATION CORRECTION (XAC)

Tomographic reconstruction of an attenuation image from a single-view mammogram projection violates Orlov's condition

[4] rather severely. In addition, mammogram images are acquired using a distribution of x-ray energies, typically 23 – 35 kVp [5]. Thus even if a tomographic image were possible, conversion of the attenuation image to the true attenuation factors at 511 keV for PET would not have a unique solution. Fortunately, several constraints are available. First is that the two main components of breast tissue are adipose and fibroglandular tissue. While the attenuation coefficients are significantly different in the photon energy range for mammography, this difference is considerably reduced in the PET energy range and thus in a narrow range (Table 1).

Table 1. Breast tissue properties [6].

	Adipose	Fibroglandular	Difference
Density (g/cm^3)	0.9500	1.0600	10%
LAC @ 20keV (1/cm)	0.5393	0.8724	38%
LAC @ 500keV (1/cm)	0.0921	0.1017	9%

LAC is the linear attenuation coefficient.

Thus estimation of the volumetric breast boundary and assignment of a constant attenuation coefficient that is the average of that for adipose and fibroglandular tissue at 511 keV will provide a close estimate of the volumetric PET attenuation image.

An additional constraint is that the geometry of the breast under mild compression is largely defined by the known separation of the compression paddle (Figure 3). In some cases the compression paddle may be slightly distorted and/or not parallel to the bottom support. However correction methods have been developed to account for these errors [e.g. 7]. If the paddle separation (d in Figure 3) is known, then the remaining information needed for estimation of the volumetric PET attenuation image is the partial attenuation length (l in Figure 3). The partial attenuation length can be estimated from the unprocessed mammogram as we have presented previously [8]. Here we evaluate the impact of the XAC method of reconstructed PET images using simulations.

III. Methods

We used the ASIM PET simulator [9] to generate noiseless emission and attenuation sinograms for a simplistic breast phantom consisting of adipose tissue that was contained in 16 x 10 x 6 cm volume. The paddle separation (d in Figure 3) was 6 cm. The breast phantom also contained 3 spherical hot spots with diameters of 5, 10, and 20 mm and with a 6:1 lesion:background ratio. Images were reconstruction analytically with STIR [10]. In addition, using a modified version of ASIM we simulated the cone-beam acquisition geometry of a mammogram using a 35 keV photon beam onto a 660 mm wide FOV with 1295 pixels (0.51 mm square pixels). The estimated volumetric attenuation image using the XAC method is compared with the true attenuation volume (Figure 4).

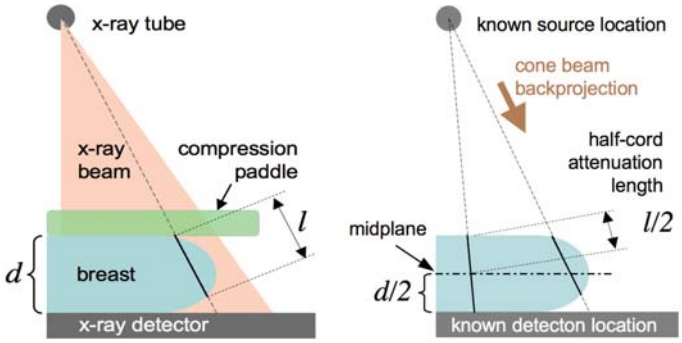
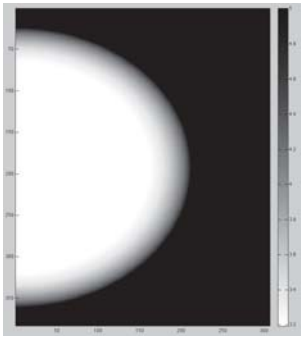
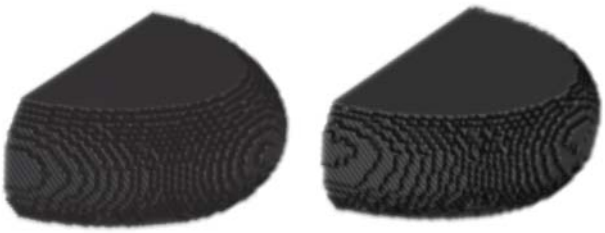


Figure 3. Illustration of the x-ray based attenuation correction method. The attenuation length in the partial-cord regions (*l*) can be directly measured from unprocessed mammograms. The attenuation length (*l*) is then symmetrically distributed about the mid-plane to estimate these boundary regions



Simulated mammogram



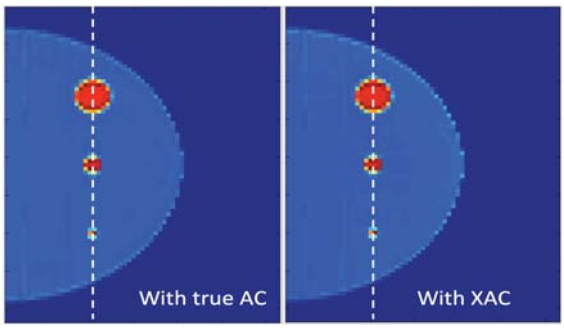
Estimated attenuation volume

Figure 4. Estimated volumetric attenuation image using the XAC method.

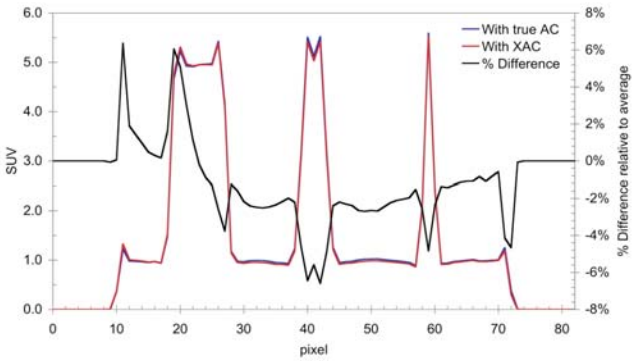
IV. Results/Discussion/Conclusion

Sections of the volumetric PET images containing the lesions are shown in Figure 5. There are small differences in the boundary locations of the volumetric attenuation images that are visible in Figure 4, but are better visualized with difference images (data not shown). These differences lead to emission errors of approximately ±6%. Reductions in these errors appear feasible by adding additional boundary information and dual-kVp and/or tomosynthesis mammogram data. The XAC method is a viable approach for attenuation correction with the

PET/X scanner. Potential challenges may arise with non-biological materials in the field of view. However additional information dual-kVp and/or tomosynthesis data can potentially be used to reduce these errors. Further investigations with more realistic data are underway.



Volumetric PET emission image sections.



Profiles (white dashed) through PET emission images.

Figure 5. Reconstructed emission images using the true AC and the XAC methods.

V. Acknowlwdgements

We thank Hannah Linden, David Mankoff, Robert Miyaoka and William Hunter. Supported by the University of Washington Coulter Foundation and NIH grant R01-CA163498.

VI. References

1. Jemal A, et al. Cancer statistics, 2010. CA Cancer J Clin 60(5):277-300, 2010. PMID: 20610543.
2. Kenny et al. European Journal of Nuclear Medicine and Molecular Imaging (2007) vol. 34 (9) pp. 1339-47.
3. MacDonald LR, Hunter WCJ, Kinahan PE, Miyaoka RS. Sensitivity-Parallax Trade-offs in the Rectangular PET/X System. Proceedings of the 2011 IEEE Nuclear Science Symposium and Medical Imaging Conference, pp: October 23-29, Valencia, Spain, 2011.
4. Orlov SS, Theory of three-dimensional reconstruction. 1. Conditions for a complete set of projections, Soviet Physics Crystallography, vol. 20, pp. 312-4, May 1975.
5. Bernhardt et al. X-ray spectrum optimization of full-field digital mammography: simulation and phantom study. Medical Physics (2006) vol. 33 (11) pp. 4337-49
6. http://www.nist.gov/pml/data/xraycoef/index.cfm
7. Heine JJ, Cao K, Thomas JA. Effective radiation attenuation calibration for breast density: compression thickness influences and correction. Biomed Eng Online 9:73, 2010. PMID: 21080916.
8. Kinahan PE, Pierce LA, Kanal, K, Macdonald LR. Method and impact of attenuation correction on quantitative accuracy for the PET/X breast imaging scanner. Journal of Nuclear Medicine, vol. 53. (abstract), 2012.
9. Comtat C, Kinahan PE, Defrise M, and Michel C, Lartizien C, Townsend DW. Simulating Whole-body PET scanning with rapid analytical methods. Proceedings of: IEEE Nuclear Science Symposium and Medical Imaging Conference, Seattle, WA, October 24-30, Vol.3 Page 1260-1264, 1999.
10. Thielemans K, Mustafovic S, Tsoumpas C. STIR: Software for Tomographic Image Reconstruction Release 2, Proceedings of: IEEE Nuclear Science Symposium and Medical Imaging Conference 2006.

A New CT Rawdata Redundancy Measure applied to Automated Misalignment Correction

Christina Debbeler, Nicole Maass, Matthias Elter, Frank Dennerlein, and Thorsten M. Buzug

***Abstract*—In computed tomography, redundantly measured rays (i.e. multiple measurements of a line integral through an object) pose a computationally efficient possibility to quantify the rawdata quality in a cost function and thus to reduce many kinds of artifacts. A general downside of such a cost function is that the portion of redundant rays is generally small and depends on the specific data acquisition geometry. The authors propose using information associated to plane integrals instead of line integrals in order to tremendously increase the rawdata utilization when formulating a cost function that quantifies the rawdata quality. In the cone–beam data acquisition geometry, which does not allow plane integrals to be measured directly, plane–information is obtained using the 2D Radon transform of the measured rawdata and a subsequent differentiation operation. The new rawdata redundancy measure is successfully applied for automatic misalignment correction.**

I. Introduction

Computed tomography (CT) is used on a daily routine in medical and industrial applications. There are powerful algorithms for 3D volume reconstruction such as the Feldkamp, Davis and Kress (FDK) algorithm [1], [2] and the Clack-Defrise algorithm [3]. Exact knowledge of the scanning device's geometrical configuration during the acquisition of each projection is inevitable for any of these algorithms. Otherwise severe misalignment artifacts like blurring and the loss of spatial resolution degrade the image quality [4].

There is numerous prior art on computed tomography misalignment correction. We focus on methods that do not require an additional CT measurement of a calibration object as it is described in reference [4]. We further do not want to place additional markers in the field of measurement [5]. Under these constraints, there are basically two ways described in the literature to do misalignment correction: Image–based [6], [7] and rawdata–based [8]. Both ways have in common, that a cost function is formulated and minimized.

In image–based methods, the cost function is formulated on the reconstructed image data (e.g. a sharpness criterion). As the image data depends on the system geometry, which is to be optimized, the computational load of each cost function evaluation includes an image reconstruction and is thus very high. On the upside, image–based methods inherently use those rawdata for misalignment correction that contribute to the final image.

Rawdata–based misalignment methods formulate a cost function that quantifies the inconsistency of rawdata redundancies, typically on the basis of redundantly measured rays. This does not include an image reconstruction and is therefore several orders of magnitude faster, however, the use of these methods depends on redundancies of a specific trajectory. For example, a circular short scan configuration shows only very few redundantly–measured rays, so that corresponding misalignment methods are not able to optimize the system geometry in combination with this trajectory. This is of special interest for the geometry correction of C–arm CT systems, which tend to exhibit mechanical instabilities resulting in reduced spatial resolution [2]. Even for other popular source trajectories, the portion of rays that are actually measured redundantly is very low and therefore the practical use of the computationally very efficient rawdata–domain misalignment methods is very restricted.

C. Debbeler and Prof. Dr. T. M. Buzug are with the Institute of Medical Engineering, University of Lübeck, Lübeck, Germany. Dr. N. Maass, Dr. M. Elter and Dr. F. Dennerlein are with the Siemens AG, Healthcare Sector, Erlangen, Germany. Corresponding author: nicole.maass@siemens.com

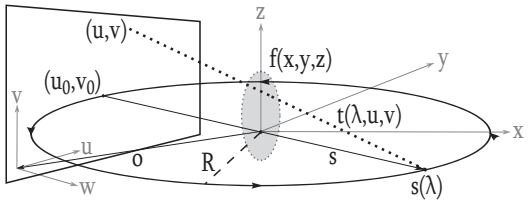


Fig. 1. Illustration of the circular 3D cone–beam acquisition geometry with a flat panel detector.

In section II we propose a method that allows using almost all acquired rawdata in a rawdata–based misalignment correction scheme, in section III we describe our experiments, which aim to proof the method, and in section IV we show our results.

II. Material & Methods

A. Definitions

Without loss of generality, we use flat detector cone–beam CT geometry throughout this work, as it is sketched in Figure 1. Cone–beam projections

$$p(\lambda, u, v) = \int_{-\infty}^{\infty} dt\, f(\mathbf{s}(\lambda) + t\mathbf{t}(\lambda, u, v)) \qquad (1)$$

of the object $f(x, y, z)$ are measured from N source positions λ_n, with $n \in [1, N]$. The scalar lambda parameterises the piecewise continuous trajectory $\mathbf{s}(\lambda)$, that contains all source positions $\mathbf{s}(\lambda_n)$, with $\lambda \in [\lambda_{\min}, \lambda_{\max}]$. The coordinates u and v parameterise the flat detector pixel's columns and rows, respectively, such that the triple (n, u, v) uniquely describes one specific ray (line integral) through the object $f(x, y, z)$.

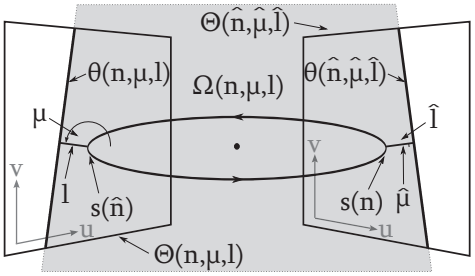


Fig. 2. Illustration of the notations used for the Radon transform and the correlation between two corresponding lines $\theta(n,\mu,l)$ and $\theta(\hat{n},\hat{\mu},\hat{l})$. The plane $\Omega(n,\mu,l)$ connecting those two lines is indicated in grey.

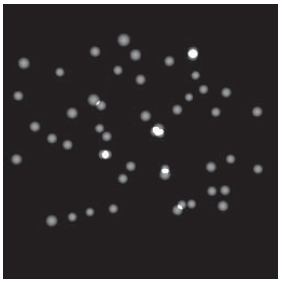


Fig. 3. Projection image of the fifty–spheres–phantom used for simulation experiments.

The right–handed world coordinate system (xyz) is fixed on the object, whereby the z–axis defines the rotation axis (if there is any). The right–handed detector coordinate system (uvw) is fixed on the detector, whereby **u**, **v**, and **w** define the unit vectors of the detector coordinate system in world coordinates. Additionally, the **o** vector defines the origin of the detector coordinate system in world coordinates. As the detector position and orientation may change in different views, these vectors (**u**, **v**, **w**, **o**) are dependent on the projection number n. The **u** (**v**) vector points from one detector column (row) to the next one and the detector indices are given as $u \in [0, U-1]$, $v \in [0, V-1]$, with U (V) being the number of detector columns (rows).

Using these definitions, we can specify a ray (n,u,v) as starting at $\mathbf{s}(\lambda_n)$ and pointing in the direction

$$\mathbf{t}(n,u,v) = \frac{\mathbf{o}(n) + u\mathbf{u}(n) + v\mathbf{v}(n) - \mathbf{s}(n)}{|\mathbf{o}(n) + u\mathbf{u}(n) + v\mathbf{v}(n) - \mathbf{s}(n)|}. \qquad (2)$$

One can see from figure 1 that the position of the detector pixel (n,u,v) in world coordinates is $\mathbf{o}(n) + u\mathbf{u}(n) + v\mathbf{v}(n)$. For the opposite direction the calculation of a projection matrix $\mathbf{P}(n)$ is necessary, which projects a world coordinate point on the detector using the geometry of view n [9] $\mathbf{P}(n) : (x,y,z) \mapsto (u,v)$.

B. Ray–based rawdata redundancy

As motivated before, we want to focus on misalignment correction methods that neither require markers nor dedicated calibration measurements and that also do not involve costly reconstruction procedures. This leads to approaches involving rawdata redundancies [8] that typically quantify and minimize the difference between redundant measurements of the same ray. Many popular source trajectories contain redundant rays

$$\mathbf{s}(\hat{n}) = \mathbf{s}(n) + t\mathbf{t}(n,u,v) \;\wedge\; \mathbf{t}(n,u,v)\cdot\mathbf{t}(\hat{n},\hat{u},\hat{v}) = 1, \quad (3)$$

with $n \neq \hat{n}$ and $t \in \mathbb{R}$. This can either be identical rays $(\mathbf{t}(n,u,v) = \mathbf{t}(\hat{n},\hat{u},\hat{v}))$ or complementary rays $(\mathbf{t}(n,u,v) = -\mathbf{t}(\hat{n},\hat{u},\hat{v}))$. According to (1), both types of redundant rays (RR) should lead to identical projection values $p(\lambda_n,u,v) = p(\lambda_{\hat{n}},\hat{u},\hat{v})$.

Consider, for instance, a full–scan circular source trajectory $\mathbf{s}_{\text{circle}}(\lambda) = (R\cos(\lambda), R\sin(\lambda), 0)^{\text{T}}$, with R being the circle radius and $\lambda_{\text{circle}} \in [0, 2\pi]$, which contains no identical rays, but numerous complementary rays in the plane of the source motion. A quantification of rawdata inconsistencies could thus involve the following cost–function

$$c_{\text{RR}} = ||p(\lambda_n,u,v) - p(\lambda_{\hat{n}},\hat{u},\hat{v})||_2 \quad \forall(n,u,v,\hat{n},\hat{u},\hat{v}), \quad (4)$$

which is to be minimized in order to correct for geometrical misalignment. Note, however, that even in case of a full–scan trajectory, only a small portion of the acquired rawdata is considered for the computation of c_{RR}. In case of short–scan or tomosynthesis geometries, only very few or even no redundant rays might be found, prohibiting the use of such a ray–based redundancy criterion.

C. Plane–based rawdata redundancy criterion

In order to establish a robust cost–function that incorporates significantly more rawdata than the approaches mentioned earlier, we propose to use redundancies in the 3D Radon domain of the object, i.e., information associated to planes. Picture any plane Ω through the object that intersects the source trajectory at least twice, at n and $\hat{n}$ with $n \neq \hat{n}$. Let us furthermore denote the intersection between Ω and the detector plane $\Theta(n) : \mathbf{x} = \mathbf{o}(n) + a\mathbf{u}(n) + b\mathbf{v}(n)$, $a,b \in \mathbb{R}$, as the line $\theta(n,\mu,l)$, which is parameterised using the detector–based quantities $\mu \in [-\pi/2, \pi/2]$ (line orientation) and $l \in [-L_{\text{max}}, L_{\text{max}}]$ (signed distance between line and detector center (u_0, v_0)) (cf. Figure 2).

In a cone–beam CT system without rawdata truncation (i.e. imaging the whole object in each projection), it is then possible to compute two redundant values associated to Ω, one involving data acquired from $\mathbf{s}(\lambda_n)$ along the line $\theta(n,\mu,l)$ and another one the data from $\mathbf{s}(\lambda_{\hat{n}})$ along the line $\theta(\hat{n},\hat{\mu},\hat{l})$.

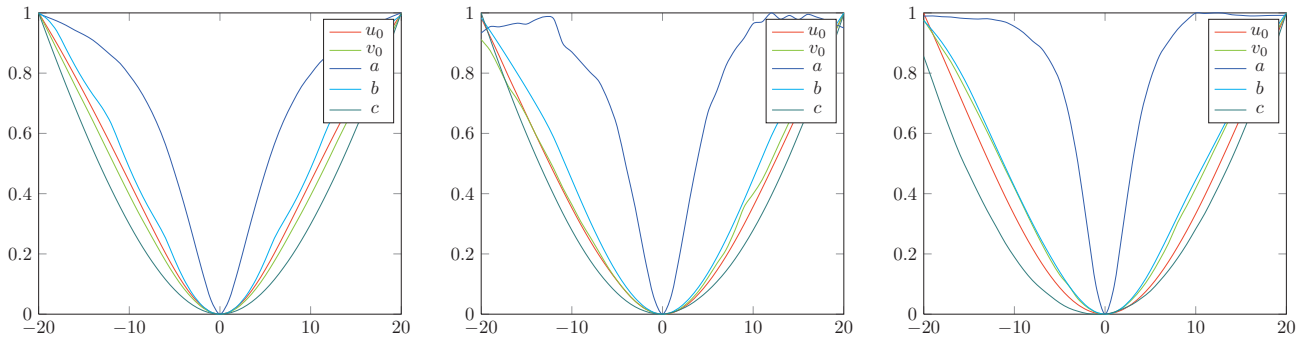
To formulate the relationship between the line data, we need to follow algorithmic steps similar to those that have been discussed in reference [3] in the context of image reconstruction. These steps operate partially in the 2D Radon domain of the rawdata, and involve the following operations:

1) Inverse cosine weighting:

$$g_1(\lambda_n,u,v) = \frac{1}{|\mathbf{w}(n)\cdot\mathbf{t}(n,u,v)|}\; p(\lambda_n,u,v) \qquad (5)$$

2) 2D Radon transform: Computation of the 2D Radon transform of the weighted projection images g_1, according to

$$g_2(\lambda_n,\mu,l) = \int_{-\infty}^{\infty} dt\, g_1(\lambda_n, l\cos\mu - t\sin\mu, l\sin\mu + t\cos\mu). \qquad (6)$$



(a) circle trajectory (b) circle+line trajectory (c) helix trajectory

Fig. 4. Relative deviation of the cost function c_{RP} from its ideal value for dejusts of selected geometry parameters (i.e. the lateral (u_0) and axial (v_0) detector offset and the detector tilt (a), slant (b), and skew (c) as defined in reference [8]). The x–axis of the plots shows a descretized number of dejust steps, whereby 41 steps are used for each geometry parameter. The step size is individually chosen for each parameter to be the smallest precision that we can reasonably achieve. This is $\Delta u_0 = 10\%$ pixel width, $\Delta v_0 = 100\%$ pixel height, $\Delta a = 0.57°$, $\Delta b = \Delta c = 0.11°$. Each curve is normalized separately to fit within a value range of 0..1 by application of an according offset and factor.

The function g_2 is thus a sinogram–like representation of the 2D projection images.

3) Differentiation: Differentiation of the sinogram–representation with respect to l.

$$g_3(\lambda_n, \mu, l) = \frac{\partial}{\partial l}\, g_2(\lambda_n, \mu, l) \tag{7}$$

This final function g_3 then contains the announced redundancies, which can now be utilized as follows:

We parameterise a plane Ω by the triple (λ, μ, l), i.e. using a detector line and a source point $\mathbf{s}(\lambda)$ to specify Ω. In the vast majority of cases, the same plane might be described using a complementary triple $(\hat{\lambda}, \hat{\mu}, \hat{l})$, with parameters depending on Ω, and therefore given in long form as

$$\begin{aligned}\hat{\lambda} &= \hat{\lambda}\left(\Omega(\lambda, \mu, l)\right),\\ \hat{\mu} &= \hat{\mu}\left(\Omega(\lambda, \mu, l)\right), \text{and}\\ \hat{l} &= \hat{l}\left(\Omega(\lambda, \mu, l)\right).\end{aligned}$$

As discussed in reference [3] we can expect $g_3(\lambda_n, \mu, l) = g_3(\hat{\lambda}, \hat{\mu}, \hat{l})$. However, while in reference [3], the authors normalize the redundancy using a proper weighting function in order to achieve accurate reconstruction results, we here intend to utilize these redundancies for misalignment quantification. We therefore introduce a cost function that quantifies data inconsistency using redundant planes (RP):

$$c_{\text{RP}} = \sqrt{\sum_{n=0}^{N-1} \sum_{\mu=-\pi/2}^{\pi/2} \sum_{l=-L_{\max}}^{L_{\max}} \left(g_3(\lambda_n, \mu, l) - g_3(\hat{\lambda}, \hat{\mu}, \hat{l})\right)^2}. \tag{8}$$

Note finally that the summation involves the fact that λ, μ, and l are discretised quantities and that all planes for which no complementary triple exists are excluded.

III. Experiments

For a general proof–of–concept we have simulated a flat detector cone–beam CT scanner with a source–detector distance of $D = 1000$ mm and a source–object distance of $R = 500$ mm. The detector has 256×256 pixel with a pitch of 2 mm and is ideally aligned such that a perpendicular through the focal spot intersects the rotation axis. We have simulated measurements using three different trajectories:

1) A full circular scan with 360 equianguarly distributed projections,
2) a partial circle plus line (CPL) scan with 216 projections on a circle segment of 216° and 114 projections on a line which is perpendicular and symmetric to the plane of rotation (line increment 8.7 mm), and
3) a helix scan with 960 projections using an angular increment of 1° and an axial increment of 0.64 mm per projection.

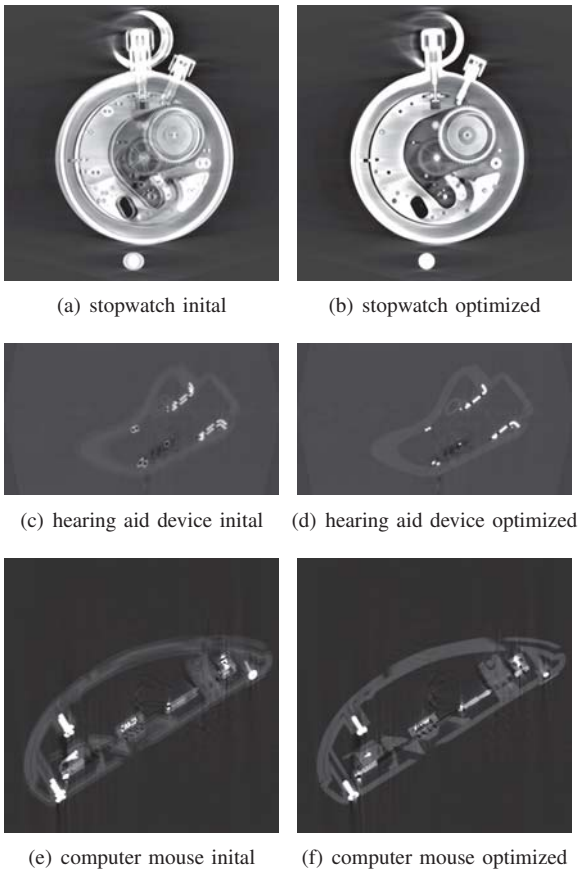
The simulated phantom consists of fifty spheres with a diameter of 10 mm that are randomly distributed within a cylinder of 256 mm diameter and 180 mm height. A projection image of the fifty–spheres–phantom is shown in figure 3. We evaluate and plot the cost function value c_{RP} as a function of dejusts of different magnitude in each trajectory parameter.

After proofing the general concept in the simulation study, the cost function has been applied to twenty datasets of different industrial cone–beam CT scanners. The datasets include a stopwatch, a hearing aid device, a metal chain, a vertebra, a computer mouse, and several electronic sensors and circuit boards. For each dataset we have started with the individual geometry that came with the dataset and varied the (from our experience) most important geometrical parameters (i.e. the lateral detector offset u_0 and the detector slant b). The cost function has been evaluated using each geometrical configuration and finally a reconstruction using the configuration with the lowest cost function value has been performed. For evaluation the reconstruction results have been visually inspected. Thereby good image quality means that there are no doubled structures, no obviously unsharp edges, and point–like objects do not appear blurred or ring–like.

IV. Results

A. Simulation study

Figure 4 shows the behavior of the cost function that we obtained from our simulation experiments. For all trajectory



(a) stopwatch inital (b) stopwatch optimized

(c) hearing aid device inital (d) hearing aid device optimized

(e) computer mouse inital (f) computer mouse optimized

Fig. 5. Selected industrial CT datasets before (left) and after (right) geometry optimization. Top row: Stopwatch; center row: hearing aid device; bottom row: computer mouse.

parameters we find a strictly monotonic decrease of the cost function value towards the optimum. The only exception is the CPL trajectory, where the cost function additionally shows local minima when applying a detector tilt (a). Note that the CPL and the helix trajectories naturally include axial trunction of the projection data. In the raw form of the cost function presented here we have not used any algorithmic steps to account for projection data truncation.

B. Industrial CT scans

We have found good image quality in 16 of the reconstructed industrial datasets while four datasets appear to suffer from misalignment after the correction with the proposed cost function. Having a closer look into the rawdata of the four cases, where the proposed cost function was not able to quantify the rawdata misalignment correctly, shows that

- in one case a very high rawdata truncation might be the reason,
- in one case the a parameter is the actual reason for the misalignment (we optimize merely u_0 and b),
- in one case the rawdata show very high jitter of the u_0 value, which needs to be optimized for each projection separately instead of globally, and
- in one case the optimization runs into a close local minimum.

Note that industrial datasets generally suffer from projection truncation in axial direction which arises from the specimen holder. In addition to the specimen holder, a significant axial data truncation (caused by the object) appeared in twelve datasets. Additionally, three of these twelve datasets have shown significant lateral data trunction.

V. Discussion and Conclusion

We proposed a new rawdata redundancy criterion that involves the 2D Radon domain of the acquired rawdata to quantify CT rawdata inconsistencies and to finally correct for geometrical misalignment.

Simulation experiments show that this cost function is able to clearly identify the correct geometry of a CT dataset under idealized simulation conditions. The simulations have also shown that the method still converges to the correct geometry if there are completely truncated projections part of the dataset. Note that the helix trajectory was configured such that the first and the last projections are completely empty.

Experiments with twenty industrial cone–beam CT datasets have shown the general applicability of the cost function to real–world problems, however, further work is necessary in order to increase its reliability. Further work will consider a method to deal with truncated projections and to optimize the projection matrix of each projection separately.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America*, vol. 1, pp. 612–619, 1984.

[2] T. M. Buzug, *Computed Tomography*. Berlin/Heidelberg, Germany: Springer, 2008.

[3] M. Defrise and R. Clack, "A cone-beam reconstruciton algorithm using shift-variant filtering and cone-beam backprojection," *IEEE TMI*, vol. 13, no. 1, pp. 186–195, March 1994.

[4] S. Sawall, M. Knaup, and M. Kachelrieß, "A robust geometry estimation method for spiral, sequential and circular cone-beam micro-ct," *Medical Physics*, vol. 39, no. 9, pp. 5384–5392, 2012. [Online]. Available: http://link.aip.org/link/?MPH/39/5384/1

[5] Y. Cho, D. J. Moseley, J. H. Siewerdsen, and D. A. Jaffray, "Accurate technique for complete geometric calibration of cone-beam computed tomography systems," *Medical Physics*, vol. 32, no. 4, pp. 968–983, April 2005. [Online]. Available: http://link.aip.org/link/?MPH/32/968/1

[6] T. Varslot, A. Kingston, G. Myers, and A. Sheppard, "High-resolution helical cone-beam micro-CT with theoretically-exact reconstruction from experimental data," *Medical Physics*, vol. 38, no. 10, pp. 5459–5476, 2011. [Online]. Available: http://link.aip.org/link/?MPH/38/5459/1

[7] A. Kingston, A. Sakellariou, T. Varslot, G. Myers, and A. Sheppard, "Reliable automatic alignment of tomographic projection data by passive auto-focus," *Medical Physics*, vol. 38, no. 9, pp. 4934–4945, 2011. [Online]. Available: http://link.aip.org/link/?MPH/38/4934/1

[8] D. Panetta, N. Belcari, A. Del Guerra, and S. Moehrs, "An optimization-based method for geometrical calibration in cone-beam CT without dedicated phantoms," *Phys. Med. Biol.*, vol. 53, pp. 3841–3861, 2008.

[9] M. Kachelrieß, M. Knaup, and O. Bockenbach, "Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware," *Medical Physics*, vol. 34, no. 4, pp. 1474–1486, 2007. [Online]. Available: http://link.aip.org/link/?MPH/34/1474/1

A Translation-based Data Acquisition Method for Industrial Computed Tomography: Experimental Results

Tobias Schön*, Theobald Fuchs†, Kilian Dremel‡, Christian Reuß‡

***Abstract*—The translation-based data acquisition method is taking advantage of the fact that variations of distance between X-ray focal spot and detector will provide both different beam angles and ray paths with respect to the object. Or, in other words: the rotational movement is substituted by one or more linear movements of the X-ray source towards the object under inspection. We present the theoretical concept of this acquisition scheme and show experimental results with real data.**

***Index Terms*—keywords**

I. Introduction

Most X-ray Computed Tomography (CT) systems which are used in the field of industrial non-destructive testing (NDT) or in medical diagnostics are based on a more or less complete rotation of the object. Alternatively, this can be achieved by rotating the detector and the X-ray source around the object. Typically, the CT systems use a fan-beam or cone-beam geometry and acquire projection data from several hundreds up to thousands of rotational directions by employing multi-row or flat-panel detectors.

However, there are numerous objects, mainly in NDT applications, which are desirable to be inspected by X-ray CT but do not allow for a rotational movement. These are for example very heavy or non portable objects which cannot be accessed from all directions, for instance, a cable channel located tightly inside the corner between the two walls of a building, see Fig. 1. Obviously, even if there is a chance to insert a flat panel detector behind the cable channel, a rotation of either the tube and detector or the object itself is impossible. In this case, the only way to obtain projections from different view angles, which is essential for the reconstruction of CT slice images, is to move the tube forward or backward towards the section of interest of the object.

Supposably, several test situations could be possible in industrial inspection and safety control for the translational CT technique. But although an application in medical diagnostics seems to be possible. A rudimental approach of our method is already proposed by Yamato and Nakahama [1] who show an application in dental radiology.

At first, this article presents the theoretically concept of translational data acquisition. Afterward, the concept is verified by real data which are acquired by an experimental setup at an industrial CT system.

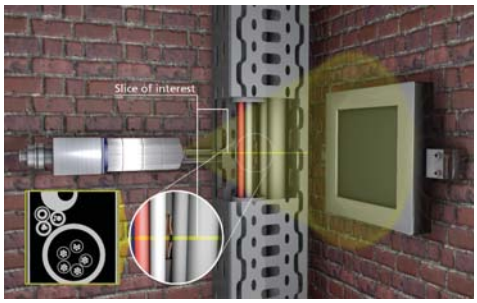


Fig. 1. A tube located close to a wall. Although it is feasible to insert a sensor into the gap between the object and the wall, the object is inaccessible to the conventional $180°$ view angle data acquisition by rotating an X-ray source and a detector.

II. Material and Methods

The translational data acquisition scheme [2], [3] is based on a linear, translational-only movement of the X-ray source. Figure 2 describes the geometry for a 2-D CT. The x-y-coordinate system is fixed with the object which is located within a circular field of measurement (FoM) with radius r_M. The origin of the x-y-coordinate system is set in the center of FoM. The positive direction of the x-axis is directed to the right and positive direction of the y-axis to the top.

Given a certain distance from the source position x_n to the center of FoM, each detector position defines a different ray angle θ with respect to the object. The t-coordinate describes the shortest distance from the ray to the center of FoM.

The ray hitting the detector at exactly $90°$ is referred as the central ray of the X-ray source. Thus, the central ray is deliberately shifted to the edge of the FoM. d_S is the maximum distance from the source to the edge of r_M. y_D is the distance from the central ray of the X-ray source to a fixed detector position. The translational movement of the source is executed parallel to the x-axis. The detector is aligned parallel to the y-axis.

In a 2-D plane each single ray can be defined by the two parameters t, θ and mathematically described by the two equations:

$$\theta = \arctan\left(\frac{y_D}{x_n + x_D}\right), \tag{1}$$

$$t = x_n \sin\theta - r_M \cos\theta, \tag{2}$$

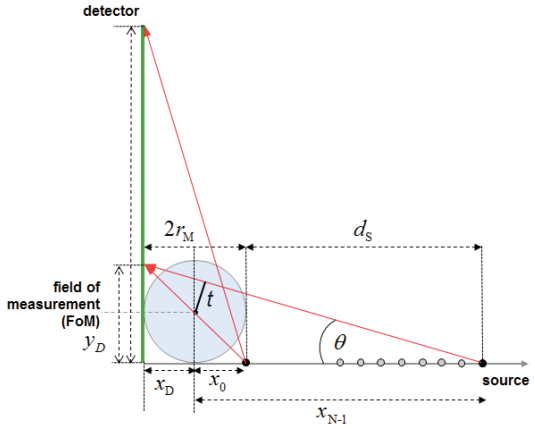


Fig. 2. Scheme of the basic principle of the translational CT data acquisition method: By changing the source position the angle θ of the ray measured at a fixed detector position changes accordingly.

where x_n $(n = 0, \cdots, N-1)$ is the distance from a certain source position to the center of the FoM and x_D is the distance from the detector to the center of the FoM. N is the number of projections acquired by a translational movement of the source.

By the two equations (1) and (2) the data range within parallel beam t-θ-coordinate space which is covered by one single translational movement can be determined (figure 3).

An equidistant sampling of source positions x_n between the source start point x_0 and stop point x_{N-1} causes a very irregular spacing of data points within t-θ-coordinate space of the the sinogram, see also [3]. Therefore, we use an heuristic approach for the sampling of the source position. $\phi = \arctan(\frac{\theta}{t})$ is the angle of a projection data-set in the t-θ-space. According to

$$\begin{aligned} \phi_{Start} &= \arctan(x_{N-1}^{-1}\,\pi/2), \\ \phi_{End} &= \arctan(x_0^{-1}\,\pi/2), \\ \Delta\phi &= \frac{1}{N-1}\left(\phi_{End} - \phi_{Start}\right), \\ x_n &= \pi/2\cot(\phi_{End} - n * \Delta\phi), \end{aligned} \tag{3}$$

with $n = 0, 1, \ldots, N-1$, a regular distribution of the data points in the t-θ-space acquired by a single translation is achieved, see figure 3.

Only a a small part of the full $180°$ parallel beam angular range can be filled by a single translational movement of the source, see figure 3. Already, in case of an infinite detector size and an infinite translational axis of the source only about one quarter of the sinogram can be filled up. Therefore, we expect that image quality is not adequate for applications in industrial CT using only one single translational movement because of the limited angular range. In order to improve the image quality we combine several translational movements (see figure 9-11, the source moves along the marked red line, the linear detector array is marked with green):

a) Translational data acquisition scheme based on two orthogonal translations along lines tangential to the FoM, which intersect behind the object (figure 9): 2T90.
b) Translational data acquisition scheme based on two translations parallel to each other (figure 10): 2Tpar .
c) A combination of 2Tpar and 2T90 (figure 11): 2(2Tpar)90.

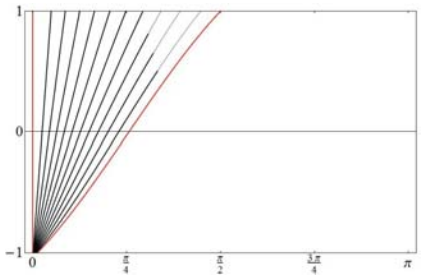


Fig. 3. The region within the parallel beam t-θ-coordinate space which can be filled by a single translational movement of the source with the non-equidistantly sampling (3). Each line depicts the data measured at one source position. The red lines show the maximum possible information which can be acquired by a detector of infinite size and no limit on the translational axis. The thick black lines show the data achieved by geometry parameters in figure 5. 10 source positions are sampled according to formula (3) along the maximum translational movement distance of the source d_S. The parameter t in (2) is normalized to $[-1, 1]$.

All measurements were acquired by the cone beam sub-μ CT system (figure 4) located at the Fraunhofer EZRT in Fürth, Germany. This system consists of an X-ray source, an X-ray detector and a manipulator system for translational and rotational movements of source, detector and object. The X-ray source is a commercially tube (FXE 160, FeinFocus from YXLON) with a transmission head. With a selected output power less than 3 W, the size of focal spot is 1.9 μm. The maximum bias voltage is 160 kV with a maximum target output power of 15 W. The X-ray detector is a commercially available flat panel sensor (C9312SK-06, Hamamatsu) with 2464×2048 pixel elements and 50×50 μm^2 pixel size.

Fig. 4. The experimental setup of the cone beam sub-μ CT system: The start position before the first projection (left) and the end position after the last projection (right). The left image shows the test object for this article, the right image shows another test object which is not used in this publication.

Figure 5 shows the geometry parameters for one translational movement of the source. The ratio detector size to FoM is about 2.5, ratio translational distance to FoM is 6. 200 source positions per translation were acquired. In all reconstruction images the non-equidistant sampling (3) was used for the source positions x_n, $n = 0, 1, \ldots, N-1$. The reconstructed matrix was 1232×1232 with a voxel size of 50 μm, so the diameter of the reconstructed slice is 6.16 cm.

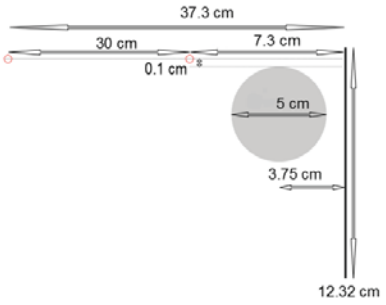


Fig. 5. Geometry parameters for a translational movement of the source: The detector size is 12.32 cm, the radius of the FoM r_M is 2.5 cm and the distance detector to the center of the FoM $x_D = 3.75$ cm. The start and stop position of the source is marked with a red circle. The source start position x_0 has a distance of 3.55 to the FoM and the maximum translational distance of the source is $d_S = 30$ cm.

The used bias voltage was 80 kV and the exposure time 500 ms for all projections. For the gain correction separated gain images without object were collected for all source positions.

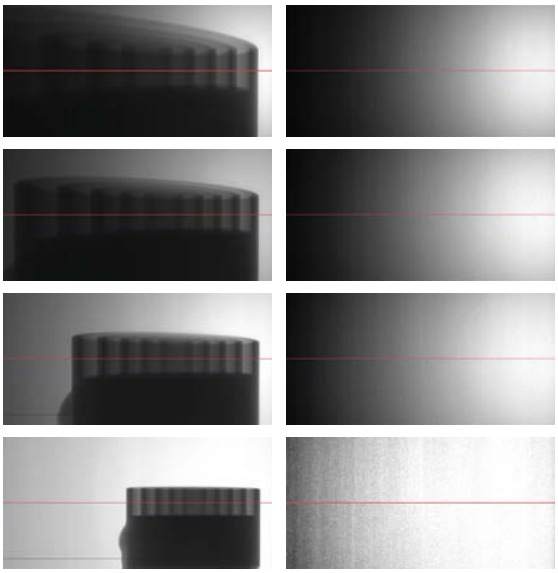


Fig. 6. The intensity image without gain correction (left) and the gain image without object (right) which is collected for all source positions and is used for the gain correction. From top to bottom: Focus detector distance is 7.3, 8.8, 10.9 and 37.3 cm.

In contrast to the classical CT acquisition geometries, such as circular or helical ct, the distance source to detector vary a lot during the data acquisition of the translational CT. In order to achieve an almost uniform unweakened intensity in all projections, the tube current was adapted during the translational movement of the source by the inverse square law. Otherwise, with the geometry parameters in figure 5 the intensity would be decrease by a factor of about 25. For the first source position the current was 3.3 μA.

Because of the completely different source to detector distances between first and last projection you recieve an unequal behavior of the measured intensity which typically does not occur in classical CT acquisition geometries. Figure

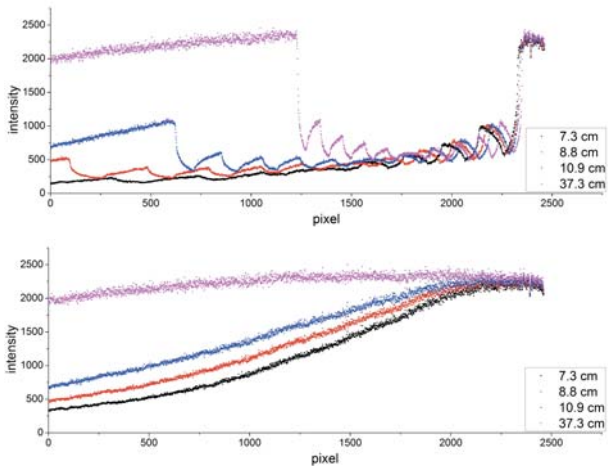


Fig. 7. Line profiles of the intensity images without gain correction (top, see left hand side of figure 6) and of the gain images without object (bottom, see right hand side of figure 6)

6 shows the intensity images without gain correction (left hand side) and gain images without object (right) for four different source positions: 7.3, 8.8, 10.9 and 37.3 cm. The line profiles of these images are depicted in figure 7. For the source positions close to the detector (see e.g. black line in figure 7) you receive a strongly decrease between the intensities measured by the central ray and the intensities measured by the ray hitting the opposite side of the detector. In contrast to this the intensities measured at the end position (see purple line in figure 7 are clearly more homogeneous.

III. Results

Out test object consists of 5 cylinders made of PVC, see figure 8 (left hand side). The sinogram of the measured projections of the central detector row acquired by the geometry 2(2Tpar)90 is visualized in figure 8 (right hand side).

Fig. 8. Left: Cylindrical test phantom consisting of PVC(left) (diameter of the outer cylinder is 5 cm, wall thickness is 2 mm and wall distance 3 mm). Right: The measured sinogram of the central detector row acquired by the geometry 2(2Tpar)90 (gain and offset corrected).

For all images the same standard SART reconstruction [4] with five iterations was applied to all data-sets and no a-prior information was utilized. All reconstruction images are displayed with the grey scale $[0.0; 1.1]$.

The simulation results in [3] illustrate that the image quality obtained by a single translation is not appropriate for typical applications in industrial CT. Therefore, we present reconstructions of real data for the three data acquisition schemes: 2T90

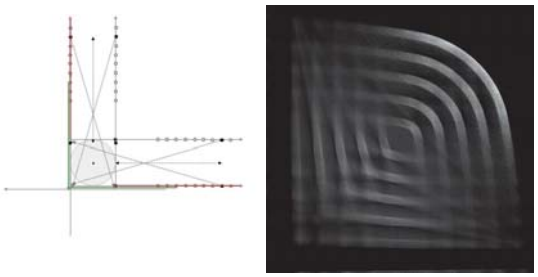


Fig. 9. Translational data acquisition scheme based on two orthogonal translations of the source (left) and the reconstruction of the central slice (right, grey scale $[0.0; 1.1]$). This method is denoted as "Two translations 90 degree" (2T90).

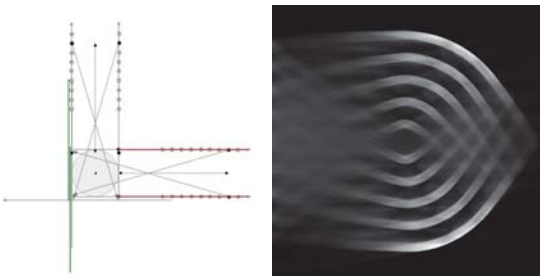


Fig. 10. Translational data acquisition scheme based on two translations parallel to each other (left) and the reconstruction of the central slice (right, grey scale $[0.0; 1.1]$. This method is called as 2Tpar.

(figure 9), 2Tpar (figure 10) and 2(2Tpar)-90 (figure 11). In all images the central slice is reconstructed. Line profiles of the central row and column with all three translational data acquisition schemes are depicted in figure 12.

In agreement with the expectations, the reconstructed images by translational CT are distorted by artifacts in a very different and locally dependent way. The quality of the resulting images depends on the position within the image and therefore noise level as well as spatial resolution may vary strongly within the same reconstructed slice, see the line profiles in figure 12.

IV. Conclusion and Future Work

We proposed results with real data of a new method for CT data acquisition, which requires almost no rotational movement of the system, respectively the object. The simulation

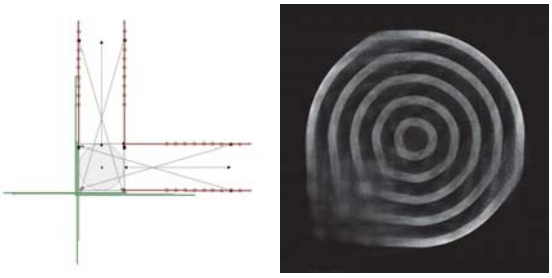


Fig. 11. A combination of 2T-par and 2T-90 (left) and the reconstruction of the central slice (right, grey scale $[0.0; 1.1]$. This acquisition scheme is denoted as 2Tpar 2(2Tpar)-90.

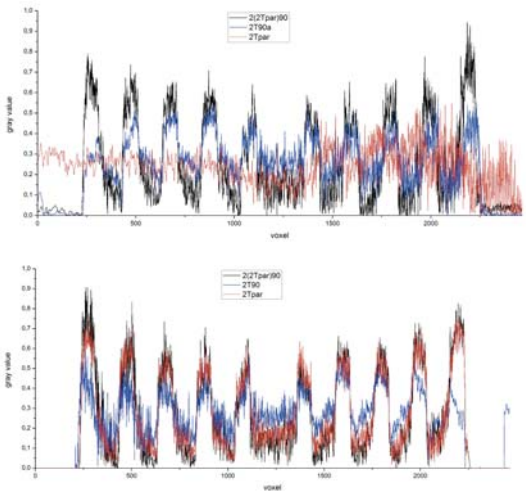


Fig. 12. Line profile of the central row (top) and central column (bottom) of the reconstructed images in figure 9, 10 and 11.

results of the Translational X-ray Computed Tomography in [3] could be verified by real data. We have shown that imaging of sections with Translational CT is feasible.

Concluding, substituting the rotational movements by linear movements of the source towards or away from the object allows for reconstructing images with adequate quality. Nevertheless, the image quality achieved by Translational CT is worse compared to a complete $180°$ parallel beam data set.

Future efforts will be made in improving the quality of the translational CT reconstructions by exploring several ways: The imaging characteristics of the translational data acquisition schemes with cone-beam geometry will be evaluated.

Real data will be acquired with a specific test specimen to analyze the image quality more thoroughly. This have to be done locally since the images with translational CT are distorted by artifacts in a very different and locally dependent way.

Last but not least, we are looking for an industrial partner to realize the concept of translational data acquisition on the basis of a real testing situation or an inspection problem.

Acknowledgment

The authors would like to thank Mike Luxa for his support during the projection data acquisition process at the cone beam sub-μ CT system.

References

[1] H. N. R. Yamato, "Study of new linear movement tomography using magnification ratio," *Journal of Meikai University School of Dentistry*, vol. 34, no. 1, pp. 39–49, 2005.

[2] T. Fuchs, T. Schön, and R. Hanke, "Translational computed tomography: A new data acquisition method for inspection of inaccessible objects," in *10th European Conference on Non-destructive Testing ECNDT, Moscow*, 2010.

[3] T. Schön, T. Fuchs, R. Hanke, and K. Dremel, "A translation-based data acquisition for industrial computed tomography," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 246–9.

[4] A. Andersen, "Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, Jan. 1984.

Coded Aperture Imaging for Fluorescent X-rays-Biomedical Applications

A. Haboub, A.A.MacDowell, S. Marchesini, D.Y.Parkinson
Lawrence Berkeley National Lab., Berkeley, California 94703, USA

Abstract— **Employing a coded aperture pattern in front of a charge couple device pixilated detector (CCD) allows for imaging of fluorescent x-rays (6-25KeV) being emitted from samples irradiated with x-rays. Coded apertures encode the angular direction of x-rays and allow for a large Numerical Aperture x-ray imaging system. The algorithm to develop the self-supported coded aperture pattern of the Non Two Holes Touching (NTHT) pattern was developed. The algorithms to reconstruct the x-ray image from the encoded pattern recorded were developed by means of modeling and confirmed by experiments. Samples were irradiated by monochromatic synchrotron x-ray radiation, and fluorescent x-rays from several different test metal samples were imaged through the newly developed coded aperture imaging system. By choice of the exciting energy the different metals were speciated.**

Index Terms— **Coded aperture, NTHT, Fluorescent x-rays**

I. INTRODUCTION

Coded aperture imaging has been used in a variety of x-ray and- γ-ray fields, such as astronomical applications, medical physics [1], plasma imaging [2], and neutron imaging [3]. It is an imaging technique without conventional lenses which was first suggested in 1961 for use in x-ray astronomy cameras [4, 5]. This imaging technique involves the use of high Z-material as the coded aperture pattern or mask, and consists of a known arrangement of transparent and opaque elements. The mask is placed between the x-ray source and a suitable sensitive detector. Transparent elements in the mask act as pinhole cameras to produce an overlapping image on the detector. This imaging technique improves the signal noise ratio (SNR) efficiency of the system over a single pinhole imaging camera.

Coded apertures encode the angular direction of x-rays, and given a known source plane, allow for a large Numerical Aperture x-ray imaging system. The algorithms to develop the self-supported coded aperture pattern of the Non Two Holes Touching (NTHT) and to reconstruct the x-ray image from the encoded pattern recorded have been developed by means of modeling and confirmed by experiments [6].

By choice of exciting x-ray radiation, fluorescent x-rays from several different test metal samples were imaged through the newly developed coded aperture imaging system. We speciate the metals by energy selecting the exciting x-rays which restrict the fluorescent x-rays available for imaging. The current CCD used lacked energy resolution, but we envisage integrating a newly developed energy resolving CCD, which will allow for imaging with elemental specificity of emitted fluorescent x-rays. Further development can include imaging sample sections with x-ray slice illumination of the sample, and by sample scanning a 3D image can be built up.

Uses for such a detector could be speciation of metals in various biological materials. Metals even with small concentration can contaminate the environment, which leads to a considerable growth of interest in trace elemental analyses and monitoring of metals in plants, animal and human tissues. For this reason several techniques have been used for monitoring these trace contaminants, such as: atomic absorption spectroscopy, graphite furnace atomic absorption spectroscopy, and x-ray fluorescence spectroscopy (XRF) [7]. This new detector has the potential to image x-ray fluorescing samples in 3D. By choice of fluorescent x-rays to be > 20 KeV for example, sample thickness can be 10's of mm so could be relevant for 3D imaging of small animals. This study represents the initial work on the detector development that can image and speciate elements in 3D within a sample.

II. THEORETICAL REVIEW AND SIMULATION

Coded aperture imaging is a two-step process. The first step, represented by equation (1), is known as the convolution. It is the generation of an overlapping image or a coded image when photons are being emitted from a source, pass through the transparent elements of the coded aperture pattern and projected onto the detector. The second step, represented by equation (2) and known as the deconvolution, is the decoding or the reconstruction of the image through a suitable reconstruction method that satisfy equation (3)

$$I(x,y) = O(x,y) \otimes A(x,y) \quad (1)$$

$$O'(x,y) = O(x,y) \otimes A(x,y) \otimes D(x,y) \quad (2)$$

$$D(x,y) \otimes A(x,y) = \delta \quad (3)$$

$I(x,y)$ represents the coded image, $O(x,y)$ is an object parallel to the coded aperture pattern whose transmission function is $A(x,y)$, and $D(x,y)$ is a decoding function or anti-mask. For an ideal imaging reconstruction, the convolution of the two

Work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

functions $D(x,y)$ and $A(x,y)$ is a system point spread function (SPSF) that is a delta function.

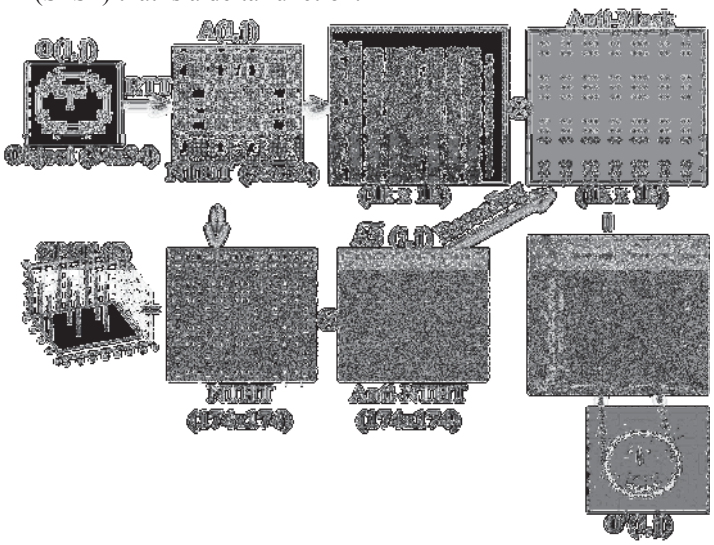


Figure 1. Coded aperture imaging simulation using ray tracing followed by a newly developed reconstruction technique that duplicates the mask and derives its anti-mask. Through the same ray tracing procedure, the anti-mask is magnified and rescaled. The image reconstruction is yielded through the Fast Fourier Transform inversion techniques.

This coded aperture imaging technique was modeled by a Ray Tracing Technique [6] by simulating the arrangement of the experimental setup that includes a source, the coded aperture pattern and detector. The simulated source is a happy face $O(i,j)$ (Fig.1) while the coded aperture pattern $A(i,j)$ 58x58 is NTHT MURA. The middle of each pixel of the source is ray traced, by passing through the middle of each pinhole of the mask and projected onto the detector. The ray tracing is followed by a newly developed reconstruction technique involving mask duplication $A3(i,j)$, anti-mask $\widetilde{A3}(i,j)$ derivation and magnification, rescaling and Fast Fourier Transform inversion techniques to yield faithful reconstruction of images $O'(i,j)$ by modeling as it is shown in Figure 1.

$$\widetilde{A3}(i,j) = \text{iFFT2}(1./\text{FFT2}(A(i,j)); \quad (4)$$

$$O'(i,j) = \text{iFFT2}(\text{FFT2}(I(i,j).* \text{FFT2}(\widetilde{B3}(i,j))) \quad (5)$$

III. EXPERIMENTS AND RESULTS

The experimental setup used to image fluorescent x-rays has been developed [6]. The coded aperture imaging system described in [6], incorporates a square 50-μm thick Tantalum mask of size 5.8 mm that is an arrangement of 421 20-μm holes in 58x58 NTHT MURA and fabricated by laser milling, as it is seen in Figure 2. The detector used is an Andor CCD detector of array size 1k x 1k with a pixel size of 13 μm. The detector dimensions are larger than those of the mask, so that this configuration provides a larger fully coded field of view (FCFV). Figure 3 illustrates the conceptual schematic of this experimental setup which is designed for 70 μm resolution over a field of view (FoV) of 6.25 mm x 6.25 mm. The sample-aperture and aperture-detector spacing are $d_1 = 70\ mm$ and $d_2 = 40\ mm$ respectively.

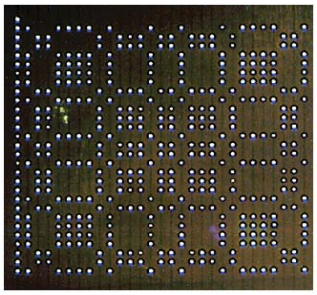


Figure 2. A magnified photograph of the 58 x 58 NTHT used in the coded aperture imaging of fluorescent x-rays. It is a square 50-μm thick tantalum mask of size 5.8 mm that has an arrangement of 421, 20-μm holes drilled by a laser miller.

The sample used in this experiment is shown in Figure 4. It consists of eight thin materials (Cr, Fe, Cu, Zn, Ge, Zr, Mo, Ag) ranging from 1-2 mm in sizes and 8 to 15 μm in thickness, and which were glued on an x-ray transparent tape. It is placed within the field of view of the system, at an angle of 45° to the x-rays direction, and at the distance d_1 from the coded aperture and $d = d_1 + d_2$ from the detector. The sample is irradiated by monochromatic synchrotron x-rays and the stimulated fluorescent x-rays that pass through the transparent positions of the coded aperture pattern are projected onto the detector, forming a coded image. These materials were chosen based on their fluorescent x-ray energies to test the capabilities of our coded aperture system for screening and imaging different fluorescent x-rays emitted from different materials at different x-ray energies.

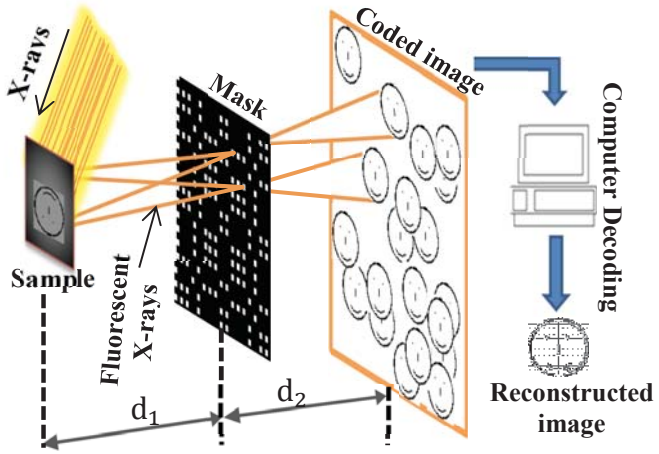


Figure 3. Conceptual schematic of the experimental setup used of coded aperture imaging for fluorescent x-rays emitted from sample when irradiated with x-rays.

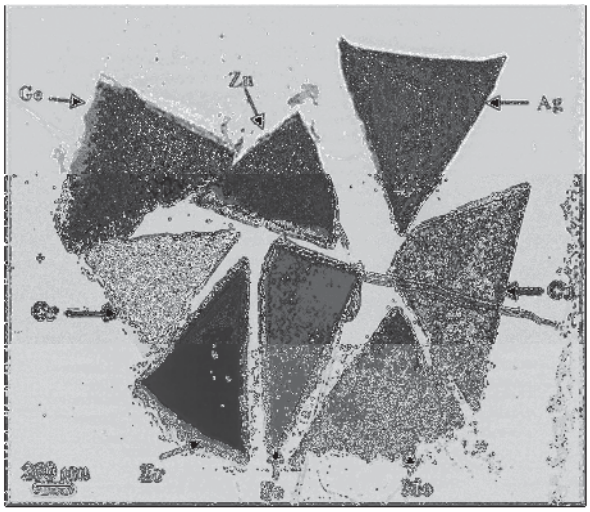


Figure 4. Radiography image of a sample, which consists of several thin metals (Cr, Fe, Cu, Zn, Ge, Zr, Mo, Ag) that emitted fluorescent x-rays (6-25KeV), when irradiated by the synchrotron radiation

At the x-ray micro-tomography Beamline 8.3.2, (Advanced Light Source, Lawrence Berkeley National Lab.) the monochromatic synchrotron x-rays illuminated the eight thin materials, which can emit a wide angular fluorescent x-rays of energies between 6 and 26 keV given adequate exciting fluorescence x-ray energy. These x-rays pass through the transparent elements of the coded aperture and form a coded image on the detector plan, as it is shown in Figure 5 for the x-ray energies of 6 keV and 26 keV respectively.

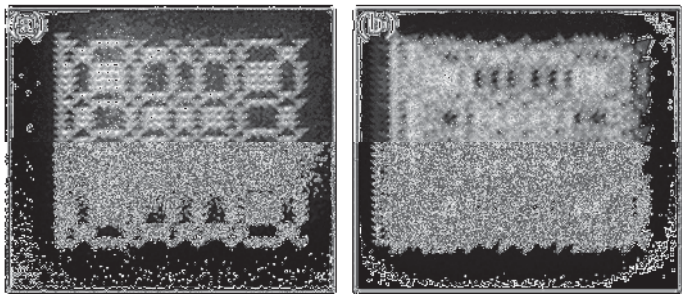


Figure 5. Example of coded images when the test materials are irradiated with x-rays of (a) energy 6 keV, (b) 26 keV.

The images were reconstructed by convoluting the coded images and its corresponding decoding functions which were obtained by the mask duplication, anti-mask derivation, magnification and rescaling process followed by the FFT inversion technique. Figure 6 shows all reconstructed images of fluorescent x-rays of energy between 6 and 25 keV that have being emitted from several materials when irradiated with x-rays radiation of different energies. Table 1 illustrated the fluorescent x-rays emitted from each material depending on the x-rays energies irradiated with. For example at x-ray energy of 6 keV, only excites the Chromium triangle to emit fluorescent x-rays, whereas at 26 keV all materials of the multi-element sample emit the fluorescent x-rays.

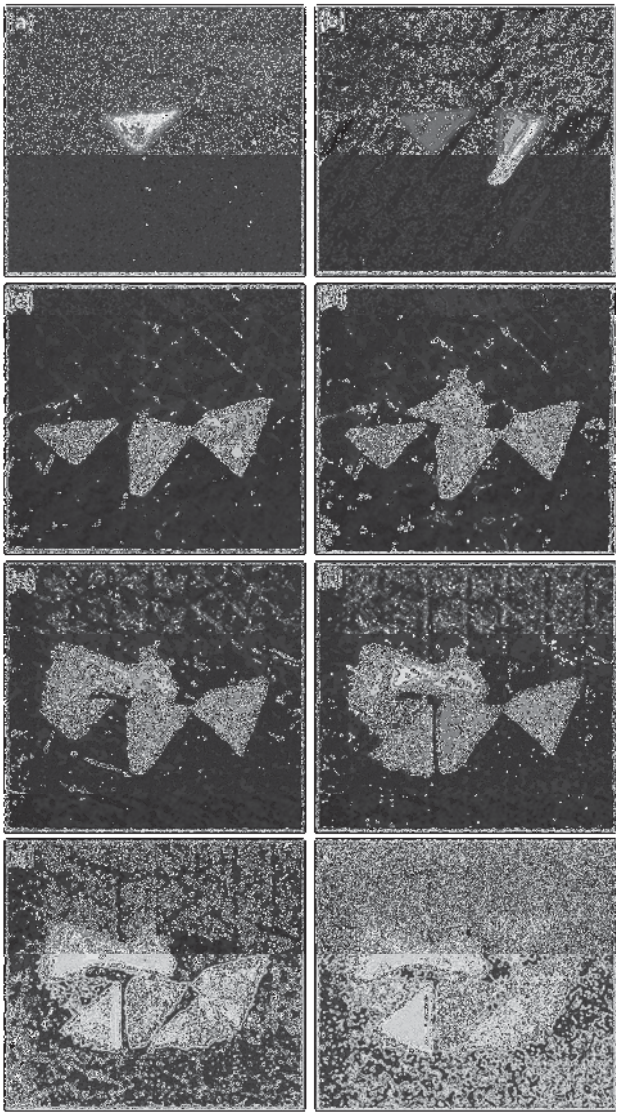


Figure 6. Reconstructed image of fluorescent x-rays (6-25KeV) that have being emitted from several materials when irradiated with x-rays radiation: (a) 6 keV (Cr) (b) 7.2 keV (Cr, Fe) (c) 9.2 keV (Cr, Fe, Cu) (d) 10 keV (Cr, Fe, Cu, and Zn) (e) 12 keV, (Cr, Fe, Cu, Zn, and Ge) (f) 19 keV, (Cr, Fe, Cu, Zn, Ge, and Zr), (g) 21 keV, (Cr, Fe, Cu, Zn, Ge, Zr, and Mo), (h) 26 keV, (Cr, Fe, Cu, Zn, Ge, Zr, Mo, Ag)

X-rays (keV)	Cr	Fe	Cu	Zn	Ge	Zr	Mo	Ag
6	x							
7.2	x	x						
9	x	x	x					
10	x	x	x	x				
12	x	x	x	x	x			
19	x	x	x	x	x	x		
21	x	x	x	x	x	x	x	
26	x	x	x	x	x	x	x	x

Table 1. Representation of fluorescent x-rays emitted from test materials as a function of the synchrotron x-rays energies

The reconstruction images are able to image all the individual metal shapes. Some reconstruction artifacts remain in the images indicating more refinement in the alignment and calibration of geometric distortion are needed. The modulation transfer function (MTF) was measured from the reconstructed images using the slanted edge method as implemented for ImageJ [8, 9].

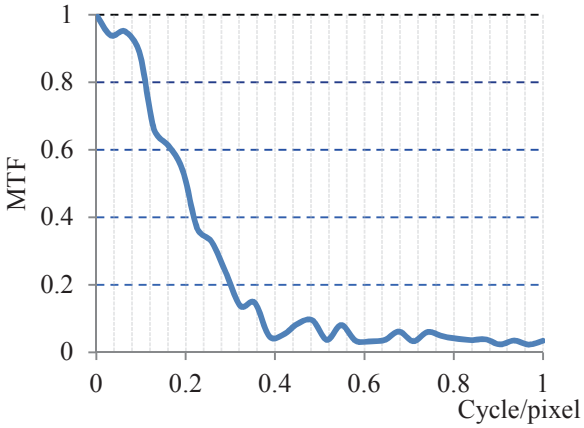


Figure 7. Modulation transfer function (MTF) as a function of cycle/pixel measured from the reconstructed image of Chromium triangle [image (a) of Figure 6.]

Figure 7 shows the MTF of the image (a) of Figure 6 as a function of the cycle/pixel. The image sharpness is closely related to the spatial frequency where MTF is 50%. In this image, the 50% of the MTF corresponds to 0.2 cycle/ pixel. The pixel size of the detector is 13 μm, thus the spatial resolution in this decoded image is about ~ 5 pixels (65 μm) at 50% MTF. This resolution is approximately equivalent to that obtained from simple geometric considerations of d_1, d_2, the size of the apertures of the mask and CCD pixel size.

IV. Future Developments

The above results indicate the initial promise of the technique. The development of the technique can include the use of an energy resolving CCD [10] to allow for elemental speciation. Improved spatial resolution can be achieved with smaller coded aperture hole size and CCD pixel size. Sensitivity can be improved by changing the current NTHT coded pattern of only 0.4% open area to a MURA pattern of 50% open area. As an example of the sensitivity of the instrument, for example in the case of irradiating a 10 μm sized contaminant of Uranium irradiated with x-rays of energy 20 keV, 80% of this energy is absorbed of which 50% is re-irradiated in the form of Lα fluorescent x-rays. Given a 50% open area, a MURA coded aperture that subtends 1 steradian and accepts 0.5 steradian. The flux density at the synchrotron is around ~10^5 hν/s/μm^2, the 10 μm thick Uranium absorbs ~8.10^6 hν, and re-irradiates ~4.10^6 hν/s into 4π steradian. Flux on the CCD is predicted to be ~1,000 hν/s. At a Uranium concentration of 0.1% the flux would reduce to 10 hν/s/ requiring several minutes to collect an image. 3D imaging can be developed by scanning an exciting X-ray slice through the sample building up the sample fluorescent images section by section. In this way a 3D element map can be imaged from within samples.

V. Conclusion

The algorithms [6] developed for the self-supported coded aperture pattern of the Non Two Holes Touching (NTHT) were used to reconstruct images of fluorescent x-rays emitted from different test materials when irradiated with x-rays at different energies. The code produced reliable images with a spatial resolution of ~ 65 μm which is approximately that derived from simple geometrical considerations. The instrument can be further developed to include elemental speciation by including an energy resolving CCD [10], improved spatial resolution by the use of smaller coded holes, improved sensitivity by the use of MURA coded apertures and 3D imaging by means of sample illumination with x-ray slices.

VI. Acknowledgement

The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] Kazachkov, Yu. P., Semenov, D. S. and. Goryacheva, N. P., "Application of Coded Apertures in Medical gamma-ray Cameras" Instrum. And Exp. Techniques 50 267-274 (2003).

[2] Chen, Y-W, Kishimoto, K.,"Tomography resolution of uniformly redundant arrays coded aperture", Rev. Sci. Instrum. 74, 2232-2235 (2003).

[3] Anderson, I. S., McGreevy, R., Bilheux, H. Z. (Eds.), "Neutron Imaging and Applications", Springer US (2009) ISBN: 978-0-387-78692-6.

[4] Mertz, L., and Young, N. O., [Proc. Int. Conf. on Opt]. Instrum. Tech., Chapman and Hall, London p.305 (1961).

[5] Caroli, E. Steven, J. B., Di Cocco, G., Natalucci, L., Spizzichino, A., "Coded aperture imaging in x-ray and Gamma-ray Astronomy", Space Science Reviews, 45, 349-403 (1987).

[6] A. Haboub ; A. A. MacDowell ; S. Marchesini ; D. Y. Parkinson; Coded aperture imaging for fluorescent x-rays. Proc. SPIE 8502, Advances in X-Ray/EUV Optics and Components VII, 850209 (October 15, 2012); doi:10.1117/12.981244.

[7] Lisa P. Colletti and George J. Havrilla, "Specimen Preparation limitations in trace element analysis quantification using micro-x-ray fluorescence", Advances in X-ray Analysis, 42 (2000).

[8] Mijta Carles, and Jaume Escofet, " Medida de la MTF en Cameras fotograficas digitales, 7th Reunion Nacional de optica, Universidad de Cantabria 2003, Santander.

[9] Boreman, Glenn D, : Modulation transfer function in optical and electro-optical sustems", SPIE Press, Bellingham, Washington.

[10] Doering, D., Chuang, Y. D., Andresen, N., Chow, K., Contarato, D., Cummings, C., Domning, E., Joseph, J., Pepper, J. S., Smith, B., Zizka, G., Ford, C., Lee, W. S., Weaver, M., atthey, L., Weizeorick, J., Hussain, Z., Denes, P., "Development of a compact fast CCD camera and resonant soft x-ray scattering end station for time-resolved pump-probe experiments", Rev. Sci. Instrum. 82, 073303 - 073303-8 (2011).

Volumetric Imaging with Sparse Arrays of Photon Counting Silicon Strip Detectors

Wojciech Zbijewski, Jennifer Xu, Steven Tilley II, J. Webster Stayman,
Katsuyuki Taguchi, Erik Fredenberg, and Jeffrey H. Siewerdsen

***Abstract*—Photon counting detectors are expected to provide numerous advantages in x-ray imaging, including elimination of electronic noise, reduction of radiation dose, and spectral imaging capabilities. An example of such high-performance detectors is edge-on silicon-strip (Si-strip) detector array introduced in scanning slot mammography systems [1]. In this paper, we present one of the first realizations of volumetric CT imaging with this detector. While application of the technology in projection mammography is well underway, its application in CT is novel and challenging from the perspectives of both system design and reconstruction algorithm development. In particular, the detector consists of a sparse arrangement of line sensors with gaps between the sensors as large as 5 mm. Orbits combining rotational and translational motions are thus necessary to achieve coverage sufficient for CT imaging. Even with such orbits, the resulting sampling pattern is non-uniform in practical imaging scenarios that seek to minimize patient exposure and scan time, necessitating a reconstruction approach beyond conventional algorithms. Iterative reconstruction approaches are known to optimally utilize all available projection data regardless of the sampling pattern and are usually more robust to under-sampling artifacts than conventional methods, pointing to such iterative methods as a natural choice for systems with sparse arrangements of radiation sensors. In light of these considerations, we have developed a bench-top photon counting CT system based on high-performance Si-strip detectors technology with capability for combined rotation-translation orbits and utilizing statistical iterative reconstruction to overcome sampling artifacts. This paper details the system design, optimization of source-detector scan orbits, performance of the reconstruction algorithm, and first image reconstructions of simple and anatomical phantoms, demonstrating the feasibility of photon-counting, volumetric CT with Si-strip edge-on detectors.**

***Index Terms*—Photon-counting Detectors, CT Reconstruction, Incomplete Sampling, Penalized-Likelihood Estimation, Extremities Imaging.**

I. Introduction

The count rates achievable by currently available photon counting x-ray detectors (PCXDs) are approaching those required for x-ray CT imaging, spurring the development of prototype and bench-top photon counting CT systems [2-5]. Among the currently available photon counting detector technologies, Silicon strip sensors have entered clinical use in commercially available scanning-slot mammography systems. Adaptation of this relatively mature and high-performance PCXD technology to CT is compelling, but a number of challenging 3D image reconstruction and system engineering problems need to be addressed. Due to the detector geometry and inability to fabricate silicon crystals of sufficiently large area coverage, the mammography detector consists of an array of line sensors separated by large gaps (up to 5 mm). While this sampling pattern is perfectly appropriate in a scanned-slot mammography system by combining data acquired during motion of the scanning arm, achieving sufficient coverage and sampling for volumetric CT requires a dedicated optimization of the scanning orbit and reconstruction algorithm to the sparse sampling pattern of the detector. Model-based, iterative reconstruction algorithms are particularly well suited to such a complex detector layout, as such methods can account for arbitrary system geometry and scanning orbit and optimally use the available projection data through appropriate definition of the system matrix involved in the projection and backprojection. Iterative reconstruction is inherently more robust to sparse scanning patterns and exotic acquisition orbits, and its potential advantages for a sparse detector matrix has been indicated in earlier studies [6]. In the current work, we present a bench-top photon counting CT system based on a Si-strip detector (Philips MicroDose, Solna, Sweden), investigate optimal acquisition orbits that minimize the effects of sparse detector sampling, and study the performance of iterative, statistical reconstruction in this imaging scenario.

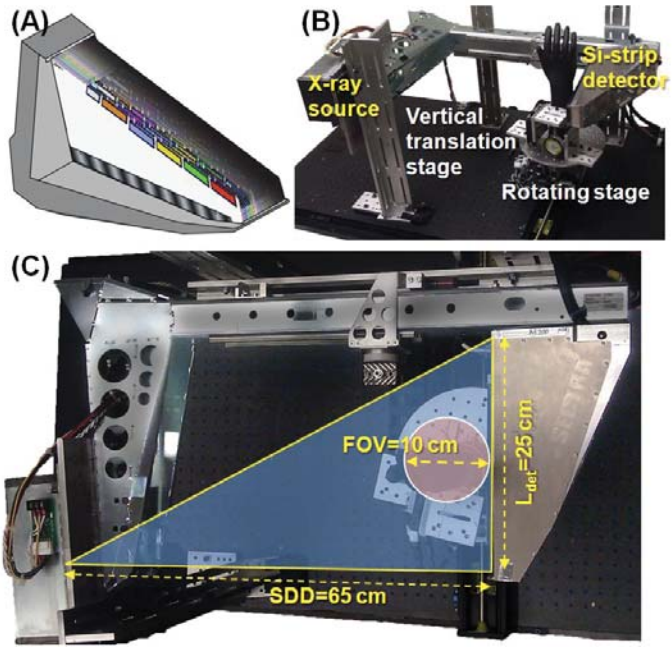


Fig. 1. (A) A photon counting Si-strip detector (developed for slot scan mammography) was used as a platform for photon-counting volumetric CT. The detector consists of an irregular, sparse array of Si-strip line sensors in edge-on configuration. The presence of significant detector gaps is challenging for conventional reconstruction algorithms and necessitates careful optimization of imaging orbits. (B) Experimental photon-counting CT bench with the Si-strip detector of (A) and a pulsed x-ray tube and motion stages allowing for combined rotation-translation scan orbits. (C) Top view of the PCXD bench with pertinent system dimensions. Note the half-fan imaging geometry.

This work was supported by NIH 2R01-CA-112163.

W. Zbijewski, J. W. Stayman, and J. H. Siewerdsen are with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21212 USA (phone: 410-955-1305; fax: 410-955-1115; e-mail: wzbijewski@jhu.edu).
K. Taguchi is with Russell H. Morgan Department of Radiology, Johns Hopkins University, Baltimore, MD 21287 USA.
E. Fredenberg is with Philips Women's Healthcare, Solna, Sweden.

II. METHODS

A. Si-strip detector

Figure 1 (A) illustrates the general design of the PCXD used in this work. The detector consists of an array of Si-strip line sensors with 50 μm pixel pitch, arranged parallel to the long axis of the detector in edge-on configuration for increased capture efficiency [1, 7]. When used for CT imaging, the detector is mounted such that the line sensors are perpendicular to the axis of rotation [Figure 1 (B) and (C)], such that the long axis of the detector is identified as the "horizontal" axis, and the short axis is identified as the "vertical" axis. The line sensors are focused on the x-ray source along both detector axes and slightly angled to avoid interactions with dead areas on sensor edges. Due to the placement of readout electronics and the aforementioned tilt of the Si-strip sensors, the resulting coverage of the detector plane is sparse, with significant dead areas between the line sensors, as illustrated in Figure 2 for a small sub-region on the detector. There are a maximum of 21 line segments along the short (vertical axis) of the detector with ~2.5 mm vertical separation between the segments. Horizontal distances between segments located in close vertical proximity are approximately 5 mm, noting that only a few line sensors present along a single horizontal slice. The total detector area is approximately 5x25 cm^2 (vertical x horizontal) with ~100,000 active pixels

B. Experimental bench and geometric calibration

The CT imaging bench developed as a platform for the Si-strip detector is shown in Figure 1 (B) and (C). A pulsed, fixed anode x-ray source (SourceRay, Ronkonkoma NY) and the photon counting detector are mounted on a rigid scanning arm to maintain geometric alignment and optimal detection efficiency with focused Si-strip detectors. The x-ray field of view of the detector in this geometry is a half-fan illustrated in Figure 1(C). The source-detector distance (SDD) is ~65 cm. The object is placed on a motorized rotating stage with the axis of rotation placed ~5 cm from the detector surface and laterally centered on the detector, resulting in ~10 cm axial Field-of-View. Vertical translational motion of the object was obtained using a manual translation stage in initial studies. A combination of vertical advancement of the object (with x-rays off) and circular scan at each vertical step (x-rays on) was used to create the various rotation-translation orbits illustrated in Figure 2 and discussed below. The source was operated at 70 kVp (+0.2 mm Cu, +2 mm Al) at 0.075 mAs per frame and 1 fps (5 mA x 15 ms pulses). The bare beam detector exposure for these operating conditions was ~0.001 mGy/pulse, corresponding to bare beam signal of ~400 counts. All acquisitions were performed at 1^o angular steps.

Geometric calibration involved precise determination of the system SDD, horizontal coordinate of the projection of perpendicular ray on the detector, and location of the center of rotation with respect to the detector. The perpendicular ray is vertically centered at the detector through mechanical construction of the scan arm. Because of the sparse detector sampling pattern, typical geometric calibration phantoms with sets of metal BBs are not well suited for this system. Instead, a phantom with two thin wires perpendicular to the axis of rotation was constructed. Sinograms of the two wires were extracted by vertically summing projection data for multiple line sensors to yield complete sampling of the wire motion. Geometric calibration parameters were found through a fit to the wire sinograms by means of a Nelder-Mead simplex algorithm. Ongoing work involves extending this approach to all geometric degrees of freedom of the system, including simultaneous determination of system geometry and exact position of each of the line sensors, which may be slightly misaligned compared to their design locations.

C. Acquisition orbits

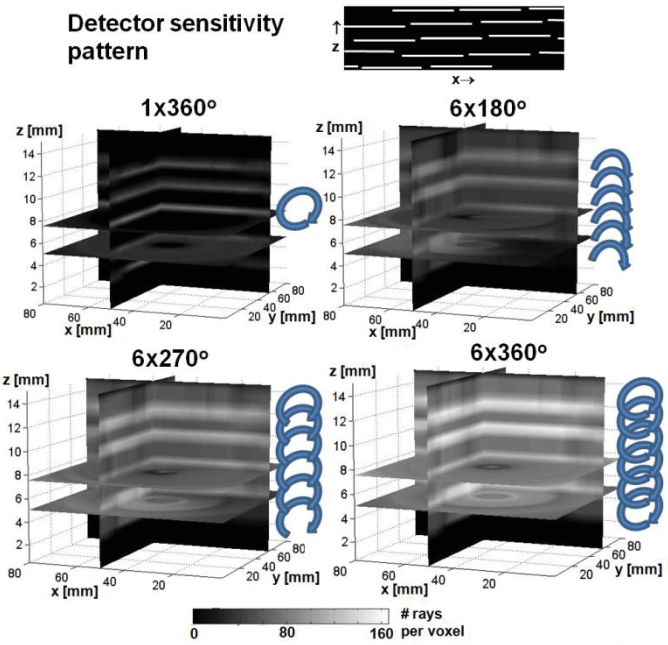


Fig. 2. Top: Detector readout (sensitivity) pattern for a small sub-region of the detector with active detector elements marked in white. Below: object sampling patterns (back-projection of the detector sensitivity pattern) over a ~80x80x15 mm^3 volume centered on the axis of rotation are shown for four scan orbits considered in this study: a single 360^o rotation, a set of six staggered 180^o rotations with 0.5 mm vertical increment, a set of six staggered 270^o rotations with 0.5 mm vertical increment, and a set of six staggered 360^o rotations with 0.5 mm vertical increment.

The extremely sparse detector sampling pattern illustrated in Figures 1 and 2 does not yield sufficient sampling of the object to enable volumetric reconstruction by conventional means. A relatively straightforward approach for improving volumetric sampling is to combine vertical motion of the object with axial circular scanning. Due to the irregular arrangement of line sensors, it is difficult to theoretically predict the optimal imaging orbit. A number of orbits have therefore been studied in simulation and investigated experimentally on the imaging bench. These include a single axial circular acquisition (denoted $1x360^o$), and three rotation-translation orbits consisting of 6 staggered axial scans separated by a 0.5 mm vertical translation of the object. These orbits differed in the range of the axial scans, including a set of 180^o scans (denoted $6x180^o$), a set of 270^o scans (denoted $6x270^o$), and a set of full circular scans (denoted $6x360^o$). While the system is ultimately likely to employ a spiral acquisition orbit, the examples considered here provide valuable insight into the complexities of object sampling and reconstruction performance for sparse detector arrays such as the Si-strip detector.

D. Iterative statistical reconstruction for sparse detector configurations

The complexities of the object sampling pattern of the Si-strip PCXD are potentially challenging for analytical reconstruction and would likely require design of a specialized

analytical inversion formula. Iterative reconstruction inherently accounts for an arbitrary system geometry through the use of an appropriate system matrix in the forward- and back-projection process, and is thus more amenable to exotic scanning geometries and sampling patterns. In this work, an iterative, statistical reconstruction (SR) algorithm was used analogous to that of Erdogan and Fessler [8]. The algorithm's objective function is based on Poisson likelihood for monoenergetic x-ray CT:

$$L(\mu; y) = \sum_i y_i \log I_{0i} e^{-l_i} - I_{0i} e^{-l_i} \qquad (1)$$

where μ is the reconstructed attenuation and y is the vector of all measured projection values (thus the index i runs through all detector pixels at all orbital steps). The projection integral through volume μ corresponding to measurement i is denoted as l_i, and I_{0i} is the bare beam signal for measurement i. Reconstruction is obtained by maximizing the following objective:

$$\hat{\mu} = \arg\max_{\mu} L(\mu; y) - \beta \cdot R(\mu) \qquad (2)$$

where $R(\mu)$ is a regularization (penalty) term. In our model, the regions of the detector corresponding to gaps in coverage are modelled as pixels with zero counts in the bare beam (I_{0i}). Those areas are thus included in the system model, simplifying the computation of the projection matrix, but are assigned a low weight by the reconstruction algorithm. The algorithm thus relies more strongly on regularization in areas corresponding to measurements that fell into the gaps.

Projection and back-projection operations needed to solve Eq. (2) were performed using an in-house GPU-implementation of the separable footprint projector of [9]. Reconstruction in each case was performed with a voxel size of 0.2x0.2x0.5 mm^3, and the reconstruction volumes were centered on the axis of rotation of the object stage. A total of 50 iterations of SR was used with 30 subsets.

III. Results

Figure 2 illustrates the object sampling pattern for each of the orbits considered in the experiments. The sampling patterns were obtained by back-projecting a mask consisting of pixels of value 1 in the regions corresponding to line sensors and of value of 0 in the gaps. Object sampling in the case of a single circular scan is highly discontinuous, and many voxels are not traveresed by any rays. This indicates the need to extened beyond a single circular scan toward a rotation-translation acquisition. While the orbits consisting of such staggered axial scans indeed show improvement in overall object sampling, a strong dependece of the sampling on the angular range of the axial scans is observed. The size of the undersampled regions of the volume diminishes with increased angular coverage, but even the 6x360^o orbit exhibits some areas of poor sampling and strong non-uniformity in the density of rays traversing the volume. The effects of sampling denisty on the quality of images obtained with SR is illustrated in Fig. 3. A hexagonal-cylindrical water phantom (~8 cm diameter) containing two tissue-mimicking inserts (bone and adipose, Gammex RMI, Madison WI) and a Cu wire was scanned. The reconstruction used a quadratic penalty with β=5, and the reconstructed volume was 80x80x15 mm^3. The two axial slices shown in Fig. 3 for each of the acquisitions correspond to the two slices through the object sampling pattern exposed in Fig. 2. For the 1x360^o orbit, the slice at z=5 mm has both a region that is partly sampled, and a central area that is not sampled. The resulting reconstruction has a smooth region in the area corresponding to the lack of sampling, where regularization fills in the gap in coverage, and a noisier region where projection data was available. For the slice at 7.5 mm and the 1x360^o acquisition, no samples were avilable, and the smooth

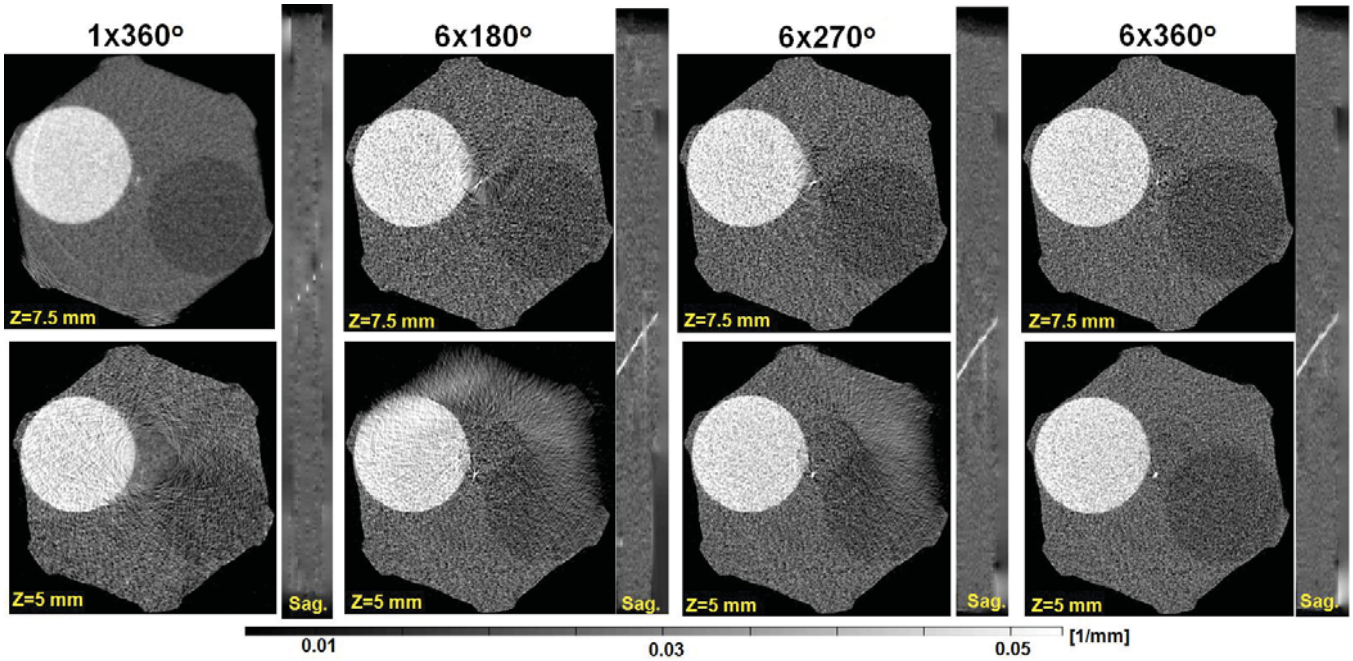


Fig. 3. Reconstructions of a water phantom with two tissue-mimicking inserts and an angled copper wire for the four orbits defined in Fig. 2. All images were obtained at 70 kVp and 0.075 mAs per frame. Two axial slices corresponding to the two slices through the object sampling pattern shown in Fig. 2 are displayed, along with a sagittal slice through the wire in the vertical strip inset on the right of each case. Regions of reduced sampling are largely filled by regularization from neighboring slices and voxels, yielding a blurry appearance in the reconstruction. Combined rotation-translation orbits involving staggered circular scans at broad angular range allow this sparse detector configuration to minimize image artifacts arising from sampling effects.

reconstruction appearance comes again from regularization filling in the gap with data from neighboring slices with better coverage. The sampling greatly improves for staggered circular scans, as observed both in the axial and sagittal slices through the reconstruction. However, even the reconstruction for the 6x270^o orbit exhibits significant areas indicative of heavy reliance on the regularizer, corresposponding to poor object sampling by the sparse detector array. There is a clear benefit from employing a 6x360^o orbit consisting of full circular scans at each vertical location.

In Fig. 4, a 3D image reconstruction of a hand phantom (consisting of cadaveric wrist and hand bones embedded in tissue-equivalent plastic) is shown for a 6x360^o acquisition. A total-variation (TV) penalty with β=0.1 was employed. The field of view was 80x80x25 mm^3. Variations in cortical thickness can be appreciated, and the trabecular structure is discernible, indicating the potential for adequate imaging performance in realistic scenarios for this early prototype system. Residual artifacts and loss of spatial resolution are attributed to - aside from the complexities of sampling patterns described above - further refinement required in geometric calibration to account for small spatial variations in detector tiling.

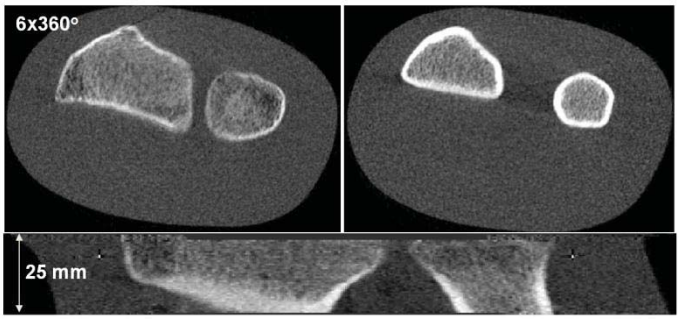


Fig. 4. Reconstructions of an anthropomorphic hand phantom with a cadaveric human skeleton in tissue-equivalent plastic. Images were reconstructed using statistical reconstruction with a total variation penalty from a set of six staggered circular scans (70 kVp, 0.075 mAs per frame). Axial and coronal slices are shown, illustrating image quality characteristics of the initial platform configuration for photon counting CT.

IV. Discussion

The results demonstrate the feasibility of volumetric CT imaging using a photon counting Si-strip detector with sparse, irregular detector sampling. The challenging sampling pattern of this detector configuration was addressed through implementation of rotation-translation acquisition orbits and the use of iterative reconstruction methods. The ability to readily incorporate arbitrary system geometries and correctly handle zero-count data (corresponding to detector gaps) in iterative reconstruction was crucial in enabling CT imaging on this platform. Artifacts associated with undersampled (or un-sampled) regions of the detector were reduced or eliminated with the iterative reconstruction approach, but images exhibited blur associated with strong regularization the gap regions. Undersampled regions in the volume resulted in a sub-optimal image appearance, mainly in the form of regions of heavy regularization. This indicates that careful design of imaging orbits to provide more complete sampling patterns is necessary, with iterative reconstruction providing the capability to readily handle even very complex scan trajectories. Ongoing work includes investigation of object sampling in spiral acquisition, orbit optimization with respect to dose and scan time, and refinement of system geometry. Detailed models of detector physics are being developed for inclusion in statistical reconstruction system models. The benchtop provides a valuable platform for investigating advanced reconstruction techniques for systems with irregularly sampled acquisition patterns as well as a test-bed for applications and performance evaluation in spectral CT.

References

[1] Fredenberg, E., Lundqvist, M., Cederstrom, B., Aslund, M., Danielsson, M., "Energy resolution of a photon-counting silicon strip detector," *Nucl. Instr. And Methods in Physics Res. A* 613:156-162 (2010).

[2] Barber, W. C., Nygard, E., Iwanczyk, J. S., Zhang, M., Frey, E. C., Tsui, B. M. W, Wessel, J. C, Malakhov, N., Wawrzyniak, G., Hartsough, N. E., Gandhi, T. Taguchi, K., "Characterization of a novel photon counting detector for clinical CT: count rate, energy resolution, and noise performance," *Proc. SPIE 7258, Physics of Medical Imaging*, (2009).

[3] Shikhaliev, P. M., "Energy-resolved computed tomography: first experimental results," *Phys. Med. Biol.* 53:5595-5613 (2008).

[4] Schlomka, J. P., Roessl, E., Dorscheid, R., Dill, S., Martens, G., Istel, T., Baumer, C., Herrmann, C., Steadman, R., Zeitler, G., Livne, A., Proksa, R., "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.* 53:4031-4047 (2008).

[5] Tomita, Y., Shirayanagi, Y. Matsui, S., Aoki, T., Hatanaka, Y., "X-ray color scanner with multiple energy discrimination capability," *Proc. SPIE* 5922, *Hard X-Ray and Gamma-Ray Detector Physics*, (2005).

[6] Zbijewski, W., Defrise, M., Viergever, M., Beekman, F. J., "Statistical reconstruction for x-ray CT systems with non-continuous detectors," *Phys. Med. Biol.* 52:403-418 (2007).

[7] Lundqvist, M., Cederstrom, B., Chmill, V., Danielsson, M., Hasegawa, B., "Evaluation of a Photon-Counting X-Ray Imaging System," *IEEE Trans. Nucl. Sci.*, 48:1530-1536 (2001).

[8] Erdogan, H., Fessler, J. A., "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.* 44:2835-51 (1999).

[9] Long, Y., Fessler, J. A., Balter, J. M., "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Im.*, 29:1839-50 (2010).

Combining Augmented Lagrangian Method with Ordered Subsets for X-Ray CT Reconstruction

Hung Nien and Jeffrey A. Fessler

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI

***Abstract*—The augmented Lagrangian (AL) method (and its closely related cousin, the alternating direction method of multipliers, or in short, ADMM) is a powerful technique for solving ill-posed inverse problems using variable splitting. In this paper, inspired by the convergence analysis of a simplified CT problem with Tikhonov regularization, we focused on the diagonal preconditioned AL method, where the step size of each entry of the split variable is proportional to the statistical weighting in the penalized weighted least squares (PWLS) formulation. To solve the inner minimization problem efficiently, we used the ordered-subsets (OS) algorithm due to its fast convergence rate in early iterations. By combining AL method with OS, experimental results show that the standard OS algorithm can be accelerated remarkably.**

I. Introduction

The augmented Lagrangian (AL) method [1] has drawn more attention recently due to its scalability, simplicity, and fast convergence property. In the field of total-variation (TV) denoising and compressed sensing, the AL method is used to split a nonsmooth term, such as the TV-norm and ℓ_1-norm, in the variational formulation, yielding a subproblem that has a closed-form solution or can be solved almost exactly [2]. In the field of statistical X-ray computed tomography (CT) image reconstruction, the AL method is also used to separate the statistical weighting matrix (which has huge dynamic range) to make the inner least squares problem much easier to precondition [3]. Aside from the standard AL method, many extensions and variations have been proposed to further accelerate convergence. A survey can be found in [4].

One variation of the AL method is to precondition the ℓ_2 penalty term in the augmented Lagrangian by some positive definite matrix $\mathbf{G}$. For example, when $\mathbf{G}$ is a diagonal matrix with positive diagonal entries, we penalize each entry in the split variable differently, which means we can have larger steps for those entries that are still far from the solution by increasing the penalty. However, such a diagonal matrix is seldom used in practice because the diagonal preconditioning matrix sometimes can ruin the opportunity to exploit fast computation such as FFT and PCG for the inner problem in the AL method.

In statistical X-ray CT image reconstruction, the image reconstruction is usually formulated as a PWLS problem, and the ordered-subsets (OS) algorithm [5] can be used to accelerate its convergence in early iterations by a factor of M, the number of subsets. This M-time acceleration comes from the "subset balance condition" by grouping the projections into M ordered subsets and updating the image incrementally using the M subset gradients. Although the standard OS algorithm approaches some limit cycle eventually because of its incremental gradient descent structure, the M-time acceleration of solving a PWLS problem is still very promising for the AL method with *inexact* updates. In this paper, we first study the convergence of a simple quadratic PWLS problem using a general AL method to get intuition about how to choose the diagonal preconditioning matrix. Then, we relax the choice of preconditioned matrix by a scaling factor, apply it to the statistical X-ray CT image reconstruction problem, and solve the inner constrained PWLS problem by using the standard OS algorithm.

II. Method

To describe our proposed algorithm more clearly, we first define the statistically weighted CT reconstruction problem as follows:

$$\hat{\mathbf{x}} \in \operatorname*{argmin}_{\mathbf{x}\in\Omega} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\}, \tag{1}$$

where $\mathbf{y}$ is the noisy post-logarithm sinogram, $\mathbf{A}$ is the system matrix of a CT scan, $\mathbf{W}$ is a diagonal weighting matrix accounting for measurement variance, R is an edge-preserving regularizer, and Ω is some convex set such as a box constraint on the solution. Instead of solving it directly using, for example, projected gradient descent method, we will focus on solving an equivalent constrained problem. That is, we are going to solve:

$$(\hat{\mathbf{x}}, \hat{\mathbf{u}}) \in \operatorname*{argmin}_{\mathbf{x}\in\Omega,\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) \right\} \text{ s.t. } \mathbf{u} = \mathbf{A}\mathbf{x}, \tag{2}$$

or equivalently, we must find a saddle point of the corresponding augmented Lagrangian of (2):

$$\mathcal{L}_A(\mathbf{x}, \mathbf{u}, \mathbf{d}) \triangleq \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{x}) + \iota_\Omega(\mathbf{x}) + \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{u} - \mathbf{d}\|_{\mathbf{G}}^2, \tag{3}$$

where ι_Ω is the characteristic function of set Ω, $\mathbf{d}$ is the scaled dual variable of $\mathbf{u}$, and $\mathbf{G}$ is some positive definite matrix, e.g., $\mathbf{G} = \eta\mathbf{I}$ with $\eta > 0$. This problem can be solved by using the alternating direction method. In other words, we will minimize $\mathcal{L}_A$ with respect to $\mathbf{x}$ and $\mathbf{u}$ alternatively followed by a gradient

ascent of $\mathbf{d}$, and the iterates will be:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \left(\mathbf{u}^{(j)} + \mathbf{d}^{(j)}\right) - \mathbf{A}\mathbf{x} \right\|_{\mathbf{G}}^2 + \mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = (\mathbf{W}+\mathbf{G})^{-1} \left(\mathbf{W}\mathbf{y} + \mathbf{G}\left(\mathbf{A}\mathbf{x}^{(j+1)} - \mathbf{d}^{(j)} \right) \right) \\ \mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} - \mathbf{A}\mathbf{x}^{(j+1)} + \mathbf{u}^{(j+1)} . \end{cases} \tag{4}$$

A. Analysis of CT problem with Tikhonov regularization

To simplify the convergence rate analysis of the proposed algorithm, we first assume that $\mathsf{R}(\mathbf{x}) = \frac{\alpha}{2}\|\mathbf{C}\mathbf{x}\|_2^2$, and Ω is the entire space, where $\mathbf{C}$ is the finite difference matrix. Then, the iterates in (4) have closed-from expressions, and by doing some simple calculations, we can show that $\mathbf{u}^{(j)}$ converges to $\mathbf{A}\left(\mathbf{A}'\mathbf{W}\mathbf{A} + \alpha\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{A}'\mathbf{W}\mathbf{y} = \mathbf{A}\hat{\mathbf{x}}$ unconditionally and linearly with rate

$$\rho\Big((\mathbf{W}+\mathbf{G})^{-1}(\mathbf{G}\mathbf{A}\mathbf{F} + \mathbf{W}) \Big) , \tag{5}$$

where $\rho(\mathbf{K})$ denotes the spectral radius of matrix $\mathbf{K}$, and

$$\mathbf{F} \triangleq \left(\mathbf{A}'\mathbf{G}\mathbf{A} + \alpha\mathbf{C}'\mathbf{C}\right)^{-1}\mathbf{A}'\left(\mathbf{G}-\mathbf{W}\right) . \tag{6}$$

Although there is no simple way to express the convergence rate in (5) using $\mathbf{G}$, one fairly good choice of $\mathbf{G}$ is $\mathbf{G} = \mathbf{W}$, thus leading to spectral radius of $1/2$, which is quite fast. However, if we set $\mathbf{G}$ to be $\mathbf{W}$, then the $\mathbf{x}$ subproblem is the original weighted CT problem with a different sinogram. In other words, the inner problem is as hard as the original problem itself, and we would gain nothing from the AL method.

B. Diagonal preconditioned AL method for CT problem

To gain something from the AL method, we must add one more degree of freedom. In this paper, we consider the preconditioning matrix $\mathbf{G} = \eta\mathbf{W}$ with $\eta > 0$, and the resulting iterates become:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \left(\mathbf{u}^{(j)} + \mathbf{d}^{(j)}\right) - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = \frac{1}{1+\eta}\left(\mathbf{y} + \eta\left(\mathbf{A}\mathbf{x}^{(j+1)} - \mathbf{d}^{(j)} \right) \right) \\ \mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} - \mathbf{A}\mathbf{x}^{(j+1)} + \mathbf{u}^{(j+1)} . \end{cases} \tag{7}$$

Intuitively, this approach penalizes the *more important* line integrals more, thus leading to *larger* step sizes for those rays. By solving the last two equations in (7), we can get the identity

$$\eta\mathbf{d}^{(j+1)} = \mathbf{y} - \mathbf{u}^{(j+1)} . \tag{8}$$

Substituting (8) into (7), the final iterates are:

$$\begin{cases} \mathbf{x}^{(j+1)} \in \underset{\mathbf{x}\in\Omega}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{z}^{(j)} - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}) \right\} \\ \mathbf{u}^{(j+1)} = \frac{1}{1+\eta}\left(\mathbf{u}^{(j)} + \eta\mathbf{A}\mathbf{x}^{(j+1)} \right) , \end{cases} \tag{9}$$

where $\mathbf{z}^{(j)} \triangleq \eta^{-1}\mathbf{y} + \left(1-\eta^{-1}\right)\mathbf{u}^{(j)}$. As can be seen from (9), the $\mathbf{x}$ subproblem is a weighted CT problem with an *updated* sinogram and a *scaled* regularizer.

To implement the proposed diagonal preconditioned AL method, we need a method to solve the inner weighted CT problem in (9). The OS algorithm is a good candidate here because it is usually fast in early iterations, and it is very easy to impose box constraints on the inner problem. Note that, when η is equal to one, the $\mathbf{x}$ subproblem is exactly the same as the original problem, and the iterates reduce to the standard OS algorithm. Intuitively, if we use a noisy FBP reconstruction as the initial guess and if we expect that the converged image should be less noisy, then we would like to choose a small η so the $\mathbf{x}$ iterate is more regularized. In general, we choose η to be between 0.3 and 1 so that we will not regularize $\mathbf{x}$ too much ($\eta < 0.3$) or too little ($\eta > 1$).

C. Practical implementation and discussion

Although (9) outlines the proposed algorithm, we usually do not implement the algorithm exactly in that way. According to the convergence theorem of ADMM methods [6, Theorem 8], it suffices for the errors of the inner minimization problems to be absolutely summable. Therefore, to try to improve the convergence behavior of our AL method, we run multiple OS iterations to refine $\mathbf{x}$ before updating the split variable $\mathbf{u}$. The practical algorithm should be as follows:

$$\begin{cases} \mathbf{x}^{(j+1)} = \mathcal{OS}_M^1\Big(\mathbf{x}^{(j)}; \frac{1}{2}\left\| \mathbf{z}^{(j)} - \mathbf{A}\mathbf{x} \right\|_{\mathbf{W}}^2 + \eta^{-1}\mathsf{R}(\mathbf{x}) , \Omega \Big) \\ \mathbf{u}^{(j+1)} = \begin{cases} \mathbf{u}^{(j)} & , \text{ if } \operatorname{mod}(j+1, P) \neq 0 \\ \frac{1}{1+\eta}\left(\mathbf{u}^{(j)} + \eta\mathbf{A}\mathbf{x}^{(j+1)} \right) & , \text{ otherwise} , \end{cases} \end{cases} \tag{10}$$

where $\mathcal{OS}_M^n(\mathbf{x}_0; \Psi, \mathcal{C})$ denotes n iterations (M sub-iterations per iteration) of the OS algorithm with initial guess $\mathbf{x}_0$, cost function Ψ, and constraint set $\mathcal{C}$, and P is the period of the split variable update. Furthermore, to minimize the error of $\mathbf{x}$ subproblem (at least for early iterations), we have to take advantage of the M-time acceleration of the OS algorithm, so the number of subsets should be large enough. However, using more subsets leads to a "larger" limit cycle, which will accelerate the error accumulation.

One could also accelerate the standard OS algorithm by starting from a larger regularization parameter (assuming the initial guess is noisy) and decreasing it gradually to one as the algorithm proceeds. The benefit of our proposed algorithm is that, thanks to the AL method, we do not have to use such continuation of the regularization parameter for convergence. Note that since the OS algorithm itself does not converge, instead of decreasing the regularization parameter (or η), we would reduce the number of subsets and increase the period of split variable update to achieve convergence in practice.

III. Result

In this section, we evaluate our proposed algorithm using a patient helical CT scan. To investigate the effects of η (the AL penalty parameter) and P (the update period), we consider three different AL penalty parameters (0.3, 0.5, and 0.7) and three different update periods (1, 5, and 10) in our experiment. The number of subsets is set to be 41. The standard OS algorithm is the baseline method. Note that each split variable update requires one "extra" forward projection compared to the standard OS algorithm. To have a fair comparison, we

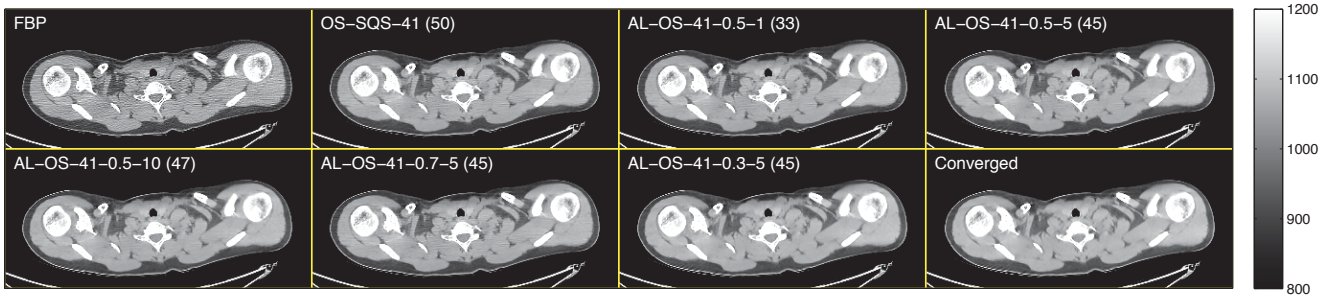


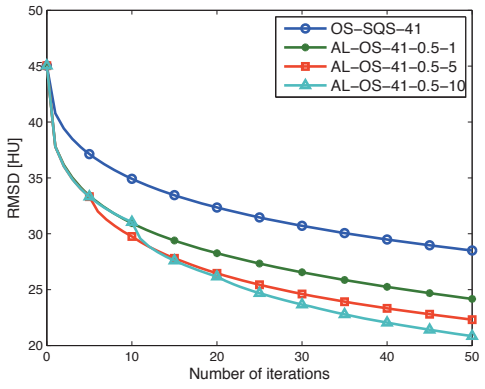
Fig. 1: Cropped images from the central slice of the reconstructed patient helical CT scan, where **FBP** denotes the FBP reconstruction, **OS-SQS-M** denotes the standard OS algorithm with M subsets, and **AL-OS-M-η-P** denotes the proposed algorithm with M subsets, the AL penalty parameter η, and the update period P. Numbers in parentheses show the number of iterations of each algorithm so that the total number of forward/back-projections is approximately 100. The **AL-OS-41-0.3-5** result after 45 iterations is very similar to the converged image, whereas the other images exhibit residual streak artifacts for the same computation time.

plot the root mean square (RMS) difference between the reconstructed image and the converged reconstruction as a function of the number of *iterations* and the number of *forward/back-projections* (assuming that $\mathbf{Ax}$ and $\mathbf{A}'\mathbf{y}$ have the same computational complexity). Lastly, since the test helical scan contains gain fluctuations [7], we include blind gain correction [8] in all of our reconstruction algorithms. With this correction, the weighting matrix $\mathbf{W}$ and the preconditioning matrix $\mathbf{G}$ are diagonal plus a rank-1 matrix rather than pure diagonal, which is a simple extension of the proposed diagonal preconditioned AL method.

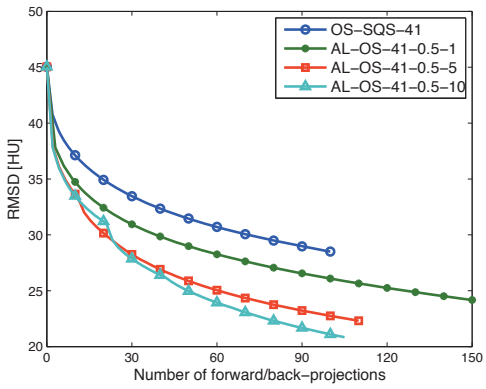
Figure 1 shows the initial noisy FBP image, the reconstructed images after about 100 forward/back-projections of the standard OS algorithm and the proposed algorithm using different values of η and P, and the converged image. As can be seen in Figure 1, the shading artifacts due to gain fluctuations are largely suppressed, and the proposed algorithm with all configurations outperforms the standard OS algorithm in image quality, especially for smaller η and larger P.

Figure 2 shows the convergence rate curves of the proposed algorithm with different values of P for the case $\eta = 0.5$, where **OS-SQS-M** denotes the standard OS algorithm with M subsets, and **AL-OS-M-η-P** denotes the proposed algorithm with M subsets, the AL penalty parameter η, and the update period P. As can be seen in Figure 2, the proposed algorithm with update period $P = 10$ converges much faster than the standard OS algorithm. There are sharp drops in the RMS difference when the split variable is updated, especially for larger P and in earlier iterations. This kind of acceleration diminishes as the algorithm proceeds because the speedup of OS algorithm saturates. To have more acceleration, we would need to either increase P or decrease M to solve the inner minimization problem in (9) more accurately.

Figure 3 shows the convergence rate curves of the proposed algorithm with different values of η for the case $P = 5$, where the naming convention is the same as in Figure 2. Note that the standard OS algorithm is just a special case of the proposed

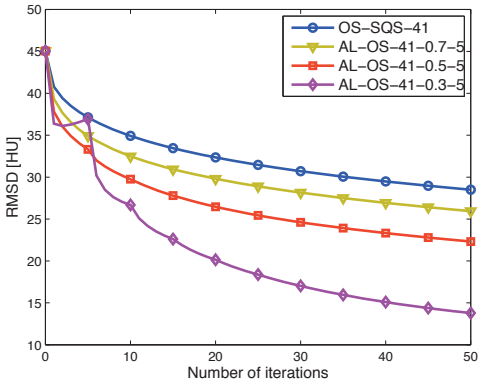


(a)

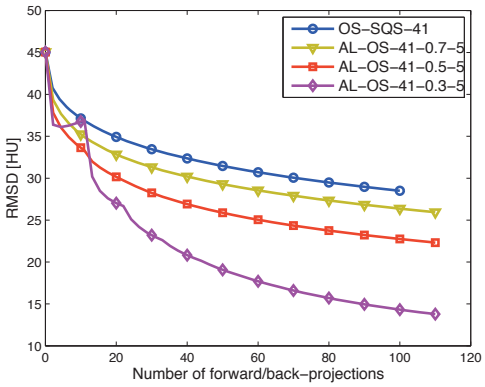


(b)

Fig. 2: RMS differences between the reconstructed image and the converged reconstruction as a function of (a) the number of iterations and (b) the number of forward/back-projections with different values of the update period P.



(a)



(b)

Fig. 3: RMS differences between the reconstructed image and the converged reconstruction as a function of (a) the number of iterations and (b) the number of forward/back-projections with different values of the AL penalty parameter η.

algorithm when $\eta = 1$. In this case, the value of P does not matter because $\mathbf{z}^{(j)}$ in (10) is independent of $\mathbf{u}^{(j)}$. As can be seen in Figure 3, the convergence rate curve converges to the curve of the standard OS algorithm as η approaches to unity. Smaller η shows faster convergence rate because the converged image is smooth and edge-preserved; however, when η is too small, for example, when $\eta = 0.3$, we can see the problem (sharp increase in RMS difference) of over-regularization in early iterations since the inner minimization problem is too different from the original problem. When the inner minimization problem is solved properly, i.e., smaller error due to larger P or M, this "misdirection" can be corrected by split variable updates, for example, the purple curves in Figure 3. Furthermore, although we consider only the standard OS algorithm in this paper, any fast variation of the OS algorithm, e.g., [9] and [10], can be applied to the proposed diagonal preconditioned AL method.

IV. Conclusion

In this paper, we proposed to combine the AL method with OS. Inspired by the convergence analysis of the AL method for quadratic PWLS problems, we focused on a diagonal preconditioning matrix $\mathbf{G}$ that is proportional to the statistical weighting matrix $\mathbf{W}$. Experimental results show that the proposed algorithm accelerated the standard OS algorithm remarkably and provides a degree of freedom to fine tune the convergence rate. As possible future work, we will investigate different splits in the proposed diagonal preconditioned AL method. In addition, we are also interested in combining, for example, the frequency analysis of the AL method with tuning the AL penalty parameter η.

Acknowledgment

This work is supported in part by NIH grant R01-HL-098686 and by an equipment donation from Intel. The authors also want to thank GE for providing the patient helical CT sinogram data.

References

[1] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, pp. 303–20, Nov. 1969.

[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Im. Proc.*, vol. 20, no. 3, pp. 681–695, 2011.

[3] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, pp. 677–88, Mar. 2012.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[5] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–51, Nov. 1999.

[6] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, Apr. 1992.

[7] J.-B. Thibault, Z. Yu, K. Sauer, C. Bouman, and J. Hsieh, "Correction of gain fluctuations in iterative tomographic image reconstruction," in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pp. 112–5, 2007.

[8] H. Nien and J. A. Fessler, "Splitting-based statistical X-ray CT image reconstruction with blind gain correction," in *Proc. SPIE 8668 Medical Imaging 2013: Phys. Med. Im.*, 2013. To appear as 8668-53.

[9] D. Kim and J. A. Fessler, "Parallelizable algorithms for X-ray CT image reconstruction with spatially non-uniform updates," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pp. 33–6, 2012.

[10] J. H. Cho and J. A. Fessler, "Accelerating ordered-subsets image reconstruction for X-ray CT using double surrogates," in *Proc. SPIE Medical Imaging 2012: Phys. Med. Im.*, p. 83131X, 2012.

Sparse shift-varying FIR preconditioners for fast volume denoising

Madison G. McGaffin and Jeffrey A. Fessler

***Abstract*—Splitting-based CT reconstruction algorithms decompose the reconstruction problem into a iterated sequence of "easier" subproblems. One relatively memory-efficient algorithm decomposes the reconstruction problem into a several subproblems, including a volume denoising problem. While easier to solve in isolation than jointly, these subproblems have highly shift-varying Hessians that are challenging to effectively precondition with circulant operators. In this work, we present an algorithm to design a positive-definite, Schatten p-norm optimal, finite impulse response (FIR) approximation to a given circulant matrix. With this algorithm, we generate efficient space-varying preconditioners for the volume denoising problem. We demonstrate that PCG with an efficient space-varying preconditioner can converge at least quickly as a split-Bregman-like algorithm while using considerably less memory.**

I. Introduction

Consider a statistical image reconstruction problem

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ J(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2 + \mathsf{R}(\mathbf{C}\mathbf{x}) \right\}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system matrix, $\mathbf{W}$ is a diagonal matrix of statistical weights, and $\mathsf{R}(\mathbf{C}\mathbf{x})$ is a convex, smooth and edge-preserving regularizer:

$$\mathsf{R}(\mathbf{C}\mathbf{x}) = \sum_{d=1}^{N_d} \beta_d \sum_{j=1}^{N} \kappa_{d,j} \phi\Big([\mathbf{C}_d \mathbf{x}]_j \Big). \quad (2)$$

The $\{\mathbf{C}_d\}_{d=1}^{N_d}$ are circulant first-order difference matrices, *e.g.*, $N_d = 13$ for 26-neighbor differences in 3D CT, and the object-dependent but constant $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$ control local regularizer strength [4]. The potential function ϕ is convex, smooth, nonnegative and even.

This minimization problem is challenging to solve directly due to the large dimension of $\mathbf{A}$, the nonlinearity of the regularizer, and the high spatial variance of the data-fit and regularizer Hessians, $\mathbf{A}'\mathbf{W}\mathbf{A}$ and $\nabla^2 \mathsf{R}(\mathbf{C}\mathbf{x})$, respectively.

Variable splitting may be used to introduce auxiliary variables to separate the terms in (1). Enforcing equality constraints between the new variables and linear functions of $\mathbf{x}$ then converts (1) into a new, equivalent, constrained minimization problem. The alternating directions methods of multipliers (ADMM) [2] may then be used to solve the new constrained optimization problem via an iterated sequence of optimization problems in each variable. This approach has the effect of splitting jointly difficult terms from one another, *e.g.*, the data-fit and regularization terms in (1). This technique has proved quite fruitful, and can handle both non-smooth regularizers and additional constraints like nonnegativity [10].

A relatively memory-efficient splitting introduces two auxiliary variables $\mathbf{u} = \mathbf{A}\mathbf{x}$ and $\mathbf{v} = \mathbf{x}$ to separate the data-fit and regularizer terms [8]. Applying ADMM to the resulting constrained optimization problem leads to an algorithm involving the following nontrivial inner optimization problems:

$$\mathbf{x}^{(j+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{\mu_u}{2} \left|\left| \mathbf{A}\mathbf{x} - \left(\mathbf{u}^{(j)} - \boldsymbol{\eta_u}^{(j)} \right) \right|\right|^2 + \frac{\mu_v}{2} \left|\left| \mathbf{x} - \left(\mathbf{v}^{(j)} - \boldsymbol{\eta_v}^{(j)} \right) \right|\right|^2, \quad (3)$$

$$\mathbf{v}^{(j+1)} = \underset{\mathbf{v}}{\operatorname{argmin}} \frac{\mu_v}{2} \left|\left| \mathbf{v} - \left(\mathbf{x}^{(j+1)} + \boldsymbol{\eta_v}^{(j)} \right) \right|\right|^2 + \mathsf{R}(\mathbf{C}\mathbf{v}), \quad (4)$$

with the scalar parameters μ_u and μ_v and the dual variables $\boldsymbol{\eta_u}$ and $\boldsymbol{\eta_v}$ introduced by the ADMM algorithm.

The ADMM does not require these subproblems to be solved exactly but only with summable absolute error taken over all iterations [3]. In practice, the ADMM algorithm will almost certainly not be run to convergence, and experience indicates that more accurate solutions to the iterated subproblems improve convergence of the algorithm as a whole. Consequently fast, even if not exact, solvers to problems (3) and (4) are desirable.

Solving (3) and (4) in isolation is "easier" than solving them jointly, but challenges in each problem remain. The tomography problem Hessian, $\mu_v \mathbf{I} + \mu_u \mathbf{A}'\mathbf{A}$, while free of shift variance induced by the statistical weights, is still far more shift-varying in cone-beam CT than in 2D, and evaluating the gradient of (3) remains very computationally expensive. While regularizer gradient evaluations are less expensive, the regularizer Hessian in (4) is highly shift-varying.

The preconditioned conjugate gradients (PCG) algorithm is an attractive candidate for both the tomography and denoising subproblems.[1] If an effective preconditioning operator $\mathbf{P} \approx \left(\nabla^2 J\right)^{-1}$ can be found, PCG converges quickly, has modest memory constraints, and updates all coordinates of the iterate simultaneously (which is attractive for high-dimensional problems and modern parallel hardware). However, designing such preconditioners can be challenging.

Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109-2122, U.S.A. Email: {mcgaffin, fessler}@umich.edu. Supported in part by NIH grant R01 HL 098686 and CPU donations by Intel. CT sinograms provided by GE Healthcare.

[1] Other rapidly converging algorithms exist for the denoising problem in particular, *e.g.*, split-Bregman-like algorithms [6]. However, these can require a prohibitive amount of memory for large reconstruction problems, *e.g.*, typical 3D helical CT problems.

Previously, the authors have preconditioned both the denoising problem and the tomography problem with circulant matrices [8]. Circulant preconditioners are attractive in part because they allow the algorithm designer to derive a shift invariant approximation of the Hessian, and immediately receive an efficient implementation of the approximation's inverse using FFTs. These features make them a good "default" preconditioner choice, but leave room for improvement by preconditioners which consider the spatial variance of the Hessian.

Because PCG with an appropriate preconditioner converges to an acceptably accurate solution using fewer gradient evaluations, we must measure the computational cost of a preconditioner relative to the computational cost of a gradient evaluation. If a preconditioner takes too much time to apply relative to a gradient evaluation, it may be more efficient to use a less computationally expensive preconditioner. For the denoising problem (4), gradient evaluations are relatively inexpensive, even compared to applying an FFT. A more computationally efficient preconditioner is desirable; sparse FIR filters may be sufficient to replace the FFT operations used to implement the conventional preconditioner.

In this work, we propose shift-varying preconditioners to tackle the denoising problem (4). In Section II we present an algorithm to design positive-definite FIR approximations to approximate a given circulant matrix. In Section III, we use these FIR filters to generate new preconditioners for the denoising problem. Results of the filter design algorithm and comparisons with other preconditioners and denoising algorithms are given in Section IV.

A. Notation

If $\mathbf{F} \in \mathbb{R}^{N \times N}$ is a circulant matrix, we use the lowercase $\mathbf{f} \in \mathbb{R}^N$ to indicate the first column or kernel of $\mathbf{F}$. The DFT of a vector $\mathbf{f}$ will be written using a hat, *e.g.*, $\text{DFT}\{\mathbf{f}\} = \widehat{\mathbf{f}}$. The vectors and matrices of all ones and zeros are written $\mathbf{1}$ and $\mathbf{0}$ respectively, whose dimension should be clear from context.

B. Schatten p-norms

If $\mathbf{F}$ is a matrix, the Schatten p-norm of $\mathbf{F}$ is the corresponding vector p-norm applied to the singular values of $\mathbf{F}$. The Schatten p-norms are unitarily invariant, so if $\mathbf{F}$ is circulant, $||\mathbf{F}||_p^p = \left|\left|\widehat{\mathbf{f}}\right|\right|_p^p$.

II. Preconditioner design

In this section, we describe an algorithm for designing a sparse, positive-definite, computationally efficient FIR filter to approximate a given circulant filter. In our experiments, the designed filters were restricted to symmetric $n \times n \times n$ blocks. One could instead use the following algorithm as an inner step of *e.g.*, the successive thinning algorithm [1] to algorithmically determine the filter footprint.

Let $\mathbf{G} \in \mathbb{R}^{N \times N}$ be a positive-definite circulant filter, and let $I \subset \{1, 2, \ldots, N\}$ indicate the desired filter footprint.

Let Ω be the set of circulant filters with the desired footprint. That is,

$$\Omega = \{\mathbf{X} : \mathbf{X} \text{ circulant and } [\mathbf{x}]_i \neq 0 \text{ only if } i \in I\}. \quad (5)$$

Our goal is to find the closest, in a Schatten p-norm sense, positive-definite filter in Ω to the given circulant matrix $\mathbf{G}$.

Satisfying both the positive-definiteness and the footprint constraints simultaneously is challenging. In fact, the positive-definiteness requirement (and the choice of any Schatten p-norm instead of the Schatten ∞-norm) distinguishes this problem from the one solved by the classical Parks-McClellan algorithm [9]. The Schatten p-norms are convex, and the set of positive-definite filters in Ω is convex, so we can use variable splitting to separate the constraints and use ADMM to solve the original problem.

Let $\mathbf{H} \in \mathbb{R}^{N \times N}$ be a circulant matrix and $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ be the augmented Lagrangian dual variable. With the equality constraint $\mathbf{H} = \mathbf{F}$, we have the following saddle point problem involving the augmented Lagrangian-like function $\mathcal{L}$:

$$\min_{\mathbf{F} \in \Omega, \mathbf{H} \succ \mathbf{0}} \max_{\mathbf{\Gamma}} \mathcal{L} = \frac{1}{2}||\mathbf{H} - \mathbf{G}||_p^p + \frac{\mu}{2}||\mathbf{H} - (\mathbf{F} + \mathbf{\Gamma})||_F^2 \quad (6)$$

$$= \frac{1}{2}\left|\left|\widehat{\mathbf{h}} - \widehat{\mathbf{g}}\right|\right|_p^p + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}} - \left(\widehat{\mathbf{f}} + \widehat{\boldsymbol{\gamma}}\right)\right|\right|_2^2, \quad (7)$$

with both $\mathbf{H}$ and $\mathbf{F}$ restricted to be circulant matrices. Equation (7) follows from the unitary invariance of the Schatten p-norms and the fact that the argument of the Frobenius norm is always a circulant matrix.[2]

Solving (7) with ADMM yields the following set of iterated updates:

$$\mathbf{f}^{(j+1)} = \text{proj}_\Omega\left(\text{IDFT}\left\{\widehat{\mathbf{h}}^{(j)} - \widehat{\boldsymbol{\gamma}}^{(j)}\right\}\right), \quad (8)$$

$$\widehat{\mathbf{h}}^{(j+1)} = \underset{\widehat{\mathbf{h}} \succ \mathbf{0}}{\text{argmin}} \ \frac{1}{2}\left|\left|\widehat{\mathbf{h}} - \widehat{\mathbf{g}}\right|\right|_p^p + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}} - \left(\widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j)}\right)\right|\right|_2^2, \quad (9)$$

$$\widehat{\boldsymbol{\gamma}}^{(j+1)} = \widehat{\boldsymbol{\gamma}}^{(j)} + \widehat{\mathbf{f}}^{(j+1)} - \widehat{\mathbf{h}}^{(j+1)}. \quad (10)$$

The $\mathbf{f}$ update requires only an FFT and zeroing of unneeded coordinates, or an IFFT-like operation which efficiently computes a small number of coefficients.

The $\mathbf{h}$ update requires that $\mathbf{H}^{(j+1)}$ be a positive-definite matrix, *i.e.*, $\widehat{\mathbf{h}} > 0$ (in practice, $\widehat{\mathbf{h}} \geq \epsilon$). For all $p \in [1, \infty)$, the update (9) is separable. In these cases, the problem can be solved unconstrained, and, because each separable problem is convex, the solution in each coordinate can then be clamped to $[\epsilon, \infty)$. If $p = \infty$, the $\mathbf{h}$ update is still convex but is nonseparable. We suggest an inexact and relatively efficient solution to the $p = \infty$ problem in Figure 1.

III. Denoising preconditioner design

We could use the filter design algorithm in Section II to replace the denoising problem's conventional circulant preconditioner with a more efficient FIR filter. In this section, we

[2] Both $\mathbf{H}$ and $\mathbf{F}$ are always restricted to be circulant matrices. The dual variable update for $\mathbf{\Gamma}$ (10) with the standard initialization $\mathbf{\Gamma}^{(0)} = \mathbf{0}$ ensures that $\mathbf{\Gamma}$ is also always circulant.

1) Let $\mathbf{d} = \max\left\{\epsilon, \widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j+1)}\right\} - \widehat{\mathbf{g}}$, and $\widehat{\mathbf{h}}(\eta) = \widehat{\mathbf{g}} + \min\{|\mathbf{d}|, \eta\} \cdot \operatorname{sign}(\mathbf{d})$.
2) Compute $\eta_{\min}$ and $\eta_{\max}$ as the extrema of the coordinates of $|\mathbf{d}|$.
3) Perform a grid search over $[\eta_{\min}, \eta_{\max}]$ using a small number of points of $J(\eta) = \frac{1}{2}\eta + \frac{\mu N}{2}\left|\left|\widehat{\mathbf{h}}(\eta) - \left(\widehat{\mathbf{f}}^{(j+1)} + \widehat{\boldsymbol{\gamma}}^{(j+1)}\right)\right|\right|_2^2$ to find η_*.
4) Return $\widehat{\mathbf{h}}(\eta_*)$.

Fig. 1: Approximate algorithm for the $\ell_2 - \ell_\infty$ $\mathbf{h}$ update (9).

instead propose a collection of FIR filters to model the inverse of the Hessian in different regions of the volume.

The Hessian of the denoising problem (4) can be written

$$\nabla^2 J_v(\mathbf{v}) = \mu_v \mathbf{I} + \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{D}_{(\mathbf{v})} \mathbf{C}_d, \tag{11}$$

where $\mathbf{D}_{(\mathbf{v})}$ is a diagonal matrix related to the $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$ and the second derivative of the potential function ϕ. At the jth voxel, we locally approximate the Hessian with a circulant filter parameterized by a voxel-dependent scalar. That is, for voxels k near j,

$$\mathbf{e}_k{}'\left[\nabla^2 J_v(\mathbf{v})\right]\mathbf{e}_j \approx \mathbf{e}_k{}'\left(\mu_v \mathbf{I} + \alpha_j \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)\mathbf{e}_j. \tag{12}$$

The scalars $\{\alpha_j\}_{j=1}^N$, which are used to approximate the variation in $\mathbf{D}_{(\mathbf{v})}$, are computed for each voxel as

$$\alpha_j = \frac{\mathbf{e}_j{}' \nabla \mathsf{R}(\varepsilon \mathbf{e}_j)}{\mathbf{e}_j{}'\left(\sum_{d=1}^{N_d} \varepsilon \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)\mathbf{e}_j}, \tag{13}$$

with $\varepsilon > 0$ small enough (on the order of 10^{-2} for an image in HU) that $\mathbf{e}_j{}' \nabla \mathsf{R}(\varepsilon \mathbf{e}_j)/\epsilon$ approximates the diagonal of the regularizer Hessian.

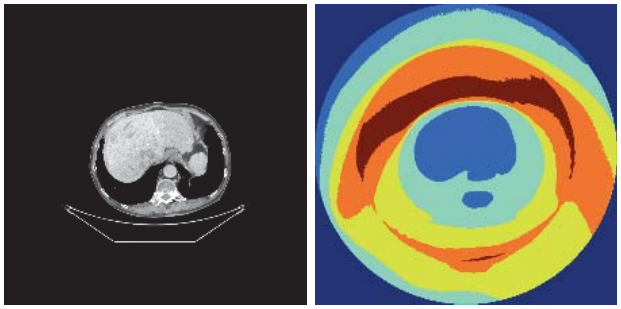
For a purely circulant filter, a single α_* is selected (*e.g.*, from the center of the volume) to form the preconditioner

$$\left(\mu_v \mathbf{I} + \alpha_* \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1}. \tag{14}$$

In this paper, we instead now quantize the $\{\alpha_j\}_{j=1}^N$ into P classes, $\{b_p\}_{p=1}^P$, using *e.g.*, the k-means algorithm. Empirically, for reasonably small P, α_j slowly varies over the volume, yielding somewhat contiguous regions with similar Hessian behavior. Motivated by this property, we propose the following preconditioner,

$$\mathbf{P}_{\text{iir}} \triangleq \sum_{p=1}^{P} \mathbf{M}_p \left(\mu_v \mathbf{I} + b_p \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1} \mathbf{M}_p, \tag{15}$$

where the $\{\mathbf{M}_p\}_{p=1}^P$ are diagonal matrices with 0 or 1 entries that partition the volume based on the voxel class assignments. Figure 2 illustrates one such partition.



(a) Image (b) Class assignments

Fig. 2: Example class assignments from the center slice of a volume with $P = 6$ classes. Note that the regions are somewhat contiguous and to a degree follow the anatomy in the volume, due to the object-dependence of the $\{\kappa_{d,j}\}_{d=1,j=1}^{N_d,N}$. Each colored region will receive a different preconditioner.

As written, (15) requires a pair of FFTs for each region in the image. This would be a significant cost for the volume denoising problem because gradient computations are relatively inexpensive. We suggest replacing the circulant inverses in (15) with a sparse FIR filter for each region designed using the algorithm described above:

$$\mathbf{P}_{\text{fir}} = \sum_{p=1}^{P} \mathbf{M}_p \mathbf{F}_p \mathbf{M}_p, \tag{16}$$

with $\mathbf{0} \prec \mathbf{F}_p \approx \left(\mu_v \mathbf{I} + b_p \sum_{d=1}^{N_d} \beta_d \mathbf{C}_d{}' \mathbf{C}_d\right)^{-1}$. This preconditioner attempts to handle the spatial variance of $\nabla^2 \mathsf{R}$, but requires no relatively expensive FFTs.

IV. Experimental results

The following experiments were performed on a helical $600 \times 600 \times 101$-voxel dataset with 888 channels, 64 rows and 2080 views provided by GE. The regularizer used 26 voxel neighbors ($N_d = 13$) and the Fair potential function with $\delta = 10.0$ Hounsfeld units:

$$\phi(t) = \delta^2(|t/\delta| - \log(1 + |t/\delta|)). \tag{17}$$

All FFTs were computed using FFTW on a 2.8 GHz Intel Core i7 CPU with 8 threads. All other operations were performed on a NVIDIA GeForce GTX 480 with 1.5 GB of global memory using OpenCL through PyOpenCL. All data was kept on the GPU unless it was necessary to page a buffer off the GPU to RAM.

We generated a 1D version of the regularizer and fit several FIR filters to a circulant approximation of its Hessian using the filter design algorithm in Section II. The results are given in Figure 3, which illustrate the trade-offs, qualitatively speaking, between different choices of Schatten p-norm and filter size. We found that the choice of Schatten p-norm did not have a significant effect on the convergence rate of PCG. Empirically, the Schatten 2-norm filter design problem seemed converge more quickly than the other choices, so we designed the filters in our next experiment with $p = 2$.

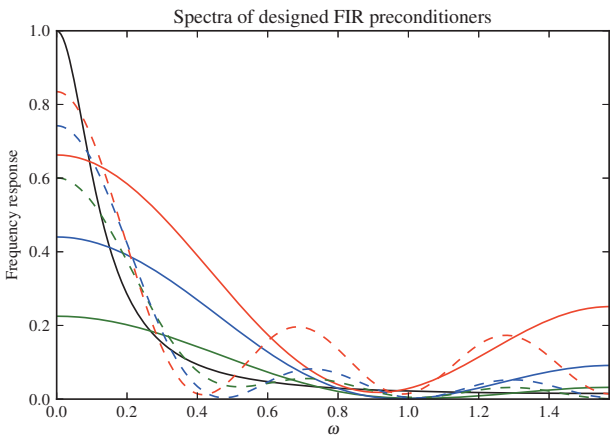


Fig. 3: Profiles of 1D filters generated by the algorithm in Section II. The target spectrum, a preconditioner for the denoising problem with $\beta = 2^4 \mu_v$, is in black. Green, blue and red correspond to the Schatten 1, 2 and ∞ norms, respectively. (These are the lower, middle and upper series on the LHS for each number of taps). The solid lines are 5-tap filters; the dashed lines are 11-tap filters.

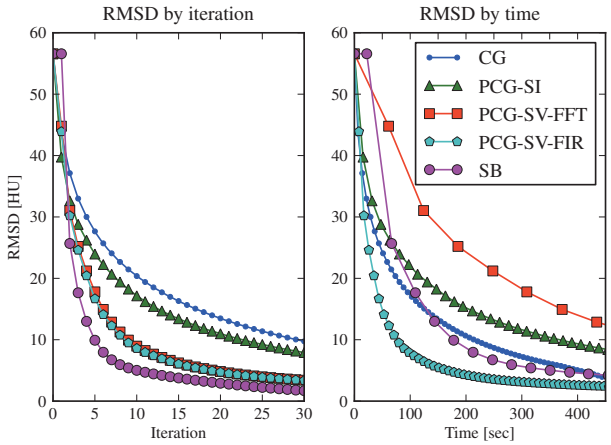


Fig. 4: Root mean-square differences (RMSD) to the converged solution of a denoising problem using (P)CG with shift-invariant (SI) and shift-varying (SV) preconditioners implemented with FFTs and $7 \times 7 \times 7$ FIR filters, and a split-Bregman (SB)-like algorithm. Six classes ($P = 6$ in (16)) were used for the shift-varying preconditioners.

We solved the denoising problem using PCG with several preconditioners and with a split-Bregman (SB) like algorithm [6]. Figure 4 shows the RMSD of each algorithm to the converged solution as a function of time and iteration.

The shift-varying preconditioner significantly outperformed the conventional circulant preconditioner, and replacing the FFTs in (15) with FIR filters had nearly no effect on per-iteration convergence rate. Remarkably, PCG with the shift-varying preconditioner converged more quickly in time than the split-Bregman algorithm. This is due in part to implementing the split-Bregman algorithm's FFTs on the CPU, which incurred GPU-CPU data transfer costs. However, to some degree these costs are unavoidable for the split-Bregman algorithm, due to the GPU's limited memory and the split Bregman algorithm's significant memory requirement. Either way, the space-varying preconditioner is a dramatic improvement over the conventional circulant filter.

V. Conclusions and Future Work

We presented an algorithm to design a positive-definite sparse FIR filter that approximates a given circulant matrix. In our experiments, we heuristically chose filters with dense cubical support, but we have no guarantee that this choice is optimal. The successive thinning algorithm [1] provides a greedy way to select the footprint algorithmically. Another possible extension is to replace the Schatten p-norm-minimization with a minimum condition number criterion, as in [7].

Sparse FIR filters can be used as part of a space-varying preconditioner to significantly accelerate the convergence of PCG applied to the volume denoising problem. The resulting algorithm is memory efficient and performs comparably to the traditionally more rapidly converging split Bregman algorithm. However, efficient preconditioners for the 3D tomography problem are still needed. Such preconditioners will likely need to account for the spatial variance of $\mathbf{A}'\mathbf{A}$, and may benefit from the locality which FIR filters provide. Early work in this direction has already been done by Fu *et. al.* [5].

Finally, while the variable splitting framework used to separate the data-fit and regularization terms has been helpful, it may be useful to revisit frameworks which combine the two. In this case, preconditioners that simultaneously account for the spatial variance present in both the data-fit and regularizer Hessians will certainly be beneficial.

References

[1] T. Baran, D. Wei, and A. V. Oppenheim. Linear programming algorithms for sparse filter design. *IEEE Trans. Sig. Proc.*, 58(3):1605–17, March 2010.

[2] D. P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12(2):133–45, March 1976.

[3] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, April 1992.

[4] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

[5] L. Fu, Z. Yu, J-B. Thibault, B. D. Man, M. G. McGaffin, and J. A. Fessler. Space-variant channelized preconditioner design for 3D iterative CT reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2013. Submitted.

[6] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–43, 2009.

[7] Z. Lu and T. K. Pong. Minimizing condition number via convex programming. *SIAM J. Matrix. Anal. Appl.*, 32(4):1193–211, 2011.

[8] M. G. McGaffin, S. Ramani, and J. A. Fessler. Reduced memory augmented Lagrangian algorithm for 3D iterative X-ray CT image reconstruction. In *Proc. SPIE 8313 Medical Imaging 2012: Phys. Med. Im.*, page 831327, 2012.

[9] T. W. Parks and J. H. McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circ. Theory*, 19(2):189–99, March 1972.

[10] S. Ramani and J. A. Fessler. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans. Med. Imag.*, 31(3):677–88, March 2012.

Effects of Sparse Sampling Schemes on Image Quality in Low-dose CT

Sajid Abbas, Taewon Lee, Hyekyun Chung, Jongduk Baek, and Seungryong Cho

***Abstract*—Various scanning methods and image reconstruction algorithms are actively investigated for low-dose CT that can potentially reduce a health risk related to radiation dose. Particularly, compressive-sensing (CS) based algorithms have been successfully developed for reconstructing images from sparsely sampled data. Although these algorithms have shown promises in low-dose CT, it has not been studied how sparse sampling schemes affect image quality in CS-based image reconstruction. In this work, we present several sparse-sampling schemes for low-dose CT, quantitatively analyze the sampling density and data incoherency, and compare effects of the sampling schemes on the image quality. We find that both sampling density and data incoherency affect the image quality, and suggest that a sampling scheme should be devised and optimized by use of these indicators.**

Index Terms—Computed tomography (CT), Compressive sensing (CS), Incoherency, Sampling density, Low-dose

I. INTRODUCTION

Computed Tomography (CT) is widely used in medical applications for its rich image information in high-quality and for its relatively short scanning time. However, multiple CT scans and/or CT scans possibly with a carelessly managed protocol in diagnostic applications including pediatric CT, or daily CT scans in other applications such as image guided radiation therapy (IGRT) may lead to a high radiation dose to the patient [1]. Low-dose CT imaging is therefore actively investigated to address this issue [2]. Radiation dose can be reduced by appropriately choosing the scanning parameters such as tube current-time product (mAs), tube voltage (kVp), section thickness, number of sections and pitch [3]. Dose reduction via controlling such scan parameters should be pursued bearing in mind a tradeoff between dose and image quality.

Recently, a compressive sensing theory that exploits the sparsity of the objective function to reconstruct from under-sampled data has been successfully translated to the CT community [4-7]. L-1 norm of a sparsifying transform of an image, e.g. magnitude of image derivative, is minimized with a constraint of data fidelity in the so-called total-variation minimization algorithm for example. The CS-based algorithms outperform the conventional analytical algorithms such as Feldkamp-Davis-Kress (FDK) algorithm in the image reconstruction problems with sparsely-sampled data. Sparsely view-sampled scheme as shown in Fig. 1(a) is one possible realization of sparse data sampling, and it has been heavily investigated in the CS frame [8-11].

However, the sparsely view-sampled scheme requires a fast on-off switching of the tube power, which may be technically challenging particularly in the diagnostic CT systems. Turning on-off of a tube in an insufficient speed with a fast gantry rotation may result in a bunched sparse-view data sampling as illustrated in Fig. 1(b). We have earlier proposed a novel method called many-view under-sampling (MVUS) that can provide a sparse data sampling without switching the tube power [12]. A multi-slit collimator is placed in between a tube and a patient, and it efficiently reduces the radiation dose to the patient. To increase the uniformity of sampling density, we proposed to reciprocate the collimator during a scan. An experimental study is also in progress and its preliminary results will be reported in another publication.

Having these various sparse sampling schemes in consideration, we were motivated to investigate the effects of sampling schemes on the image quality. Sparse data sampling schemes have not been addressed before in the context of CS-based image reconstruction in CT, and we present several sparse-sampling schemes, quantitatively analyze the sampling density and data incoherency, and compare effects of the sampling schemes on the image quality systematically in this work.

II. METHODS

In order to analyze the data property with respect to the CS-based reconstruction, we recruited two measures for each pixel: sampling density and data incoherency. Sampling density refers to the number of rays that pass through a given pixel in the image space. For simplicity and without loss of generality, we used a 2-D fan-beam CT geometry in this study. It is straightforward to implement for a 3-D cone-beam CT geometry. We first discretized the image space into a 2-D square array with 128 by 128 pixels. A 1-D detector array having 200 detector pixels was placed at 50 cm in a conventional geometry. A ray-driven approach was adopted to calculate the sampling density, i.e. the number of rays that pass through an image pixel, and the intersection lengths of the rays

The work was supported in part by the NRF grant N01100169, and by the MEST grant R0001270 and R0001376 in Korea.

Sajid Abbas, Taewon Lee, and Hyekyun Chung are with the Nuclear and Quantum Engineering Department, Korea Advance Institute of Science and Technology (KAIST), Daejeon, Korea.
Jongduk Baek is with the Yonsei Institute of Convergence Technology, Incheon, Korea.
Seungryong Cho is with Nuclear and Quantum Engineering Department, KAIST, Daejeon, Korea (telephone: +82-042-350-3828, e-mail: scho@kaist.ac.kr).

with the image pixel were accumulated for each image pixel [13].

The success of CS-approach is largely dependent upon the incoherency of the measured data. Even though the sampling density of a given pixel is high, the CS-based image reconstruction in the pixel would not be successful if all the measured data are exactly redundant for the worst example. The imaging model can be represented as below:

$$\vec{y} = \boldsymbol{A}\vec{x}, \qquad \text{Eq. (1)}$$

where $\vec{y}$ represents a vector of size *M* corresponding to the total number of ray integrals, $\vec{x}$ is a vectorized image of size *N*, and $\boldsymbol{A}$ represents the *N by M* system matrix that describes each ray integrals. Specifically, a row vector of $\boldsymbol{A}$, $\vec{a}$ represents a single ray integral. Let us suppose that the total number of rays that pass through the k^{th} pixel is *m* and that the corresponding row vectors are re-indexed from 1 to *m*. We consider all the possible combinations of two vectors out of the *m* vectors, and calculate their coherency by use of the inner product. We define data incoherency (DIC) of the k^{th} image pixel as shown in Eq. (2).

$$\mathrm{DIC_k} = 1 - \frac{\sum_{i=1}^{m}\sum_{j=i+1}^{m}\left|\frac{\overrightarrow{a_i}\cdot\overrightarrow{a_j}}{|\overrightarrow{a_i}||\overrightarrow{a_j}|}\right|}{\mathrm{m(m-1)/2}} \qquad \text{Eq. (2)}$$

The value of DIC ranges from 0 to 1, and it means that all the measurement vectors through the k^{th} pixel are orthogonal to each other if its value is 1. If the DIC value is 0, it means that all the measurements are identical.

The sampling density and the DIC are subject to sampling schemes, and we present five different sampling schemes in this work. All the sampling schemes are assumed to achieve pretty much the same reduction of radiation dose, i.e. 75 % reduction in this work, compared to a conventional scan. We divided the sampling schemes into two broad categories: (a) Few-view sampling and (b) Many-view under-sampling (MVUS). Few-view sampling is further divided into (1) sparse-view and (2) bunched sparse-view samplings. MVUS is divided into three types: (1) MVUS-Fine, (2) MVUS-Bulk, and (3) MVUS-Moving. Schematics of these sampling schemes are shown in Fig. 1. In each case, we used only 1/4 of the fully sampled data for image reconstruction. We will discuss all of the schemes in detail in the following.

A. Few-View Sampling

A direct and intuitive approach to sparse data sampling is to use fewer projection views than conventional. Few-view sampling is classified into two types.

1) Sparse-View Sampling

In the sparse-view sampling scheme, we can scan the object at regular intervals in projection angle. This work assumes that every fourth view data is used in the image reconstruction. In other words, we selected 1/4 of the data which amount to 90 views out of 360 views in this study.

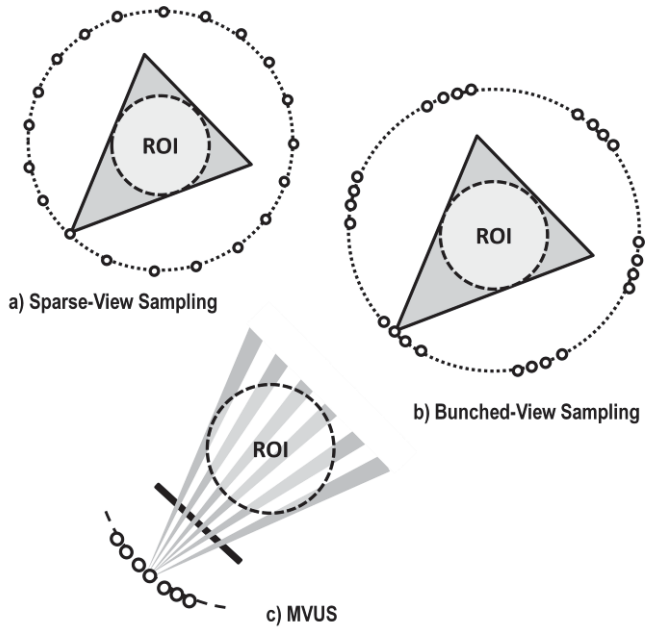


Figure 1: Schematic illustration of spare sampling schemes.

2) Bunched Sparse-View Sampling

The total number of projection views used in this sampling scheme is the same as is used in the sparse-view sampling, but the positions of and the intervals between them are different. 10 bunches, each of which consists of 9 consecutive projection views, were used in this study. An interval corresponding to 27 consecutive views was introduced between each of the two bunches.

B. Many-view Under-sampling (MVUS)

The MVUS method samples the data by use of a multi-slit collimator resulting in a sparse sampling in the detector bins for all the projection angles. MVUS is categorized into three types; a) MVUS-Fine, (b) MVUS-Bulk and c) MVUS-Moving.

1) MVUS-Fine

In this sampling scheme, we used the data from each fourth bin in a detector row. In contrast to the few-view sampling approach, projections of an object are acquired from all the 360 views. Although this sampling scheme may not be amenable to bring into practice with the current technology, but for the sake of understanding and designing the sampling protocol we investigated this scheme as well.

2) MVUS-Bulk

In this scheme, the sampled detector bins were bunched. Six bunches were placed in a regular pattern with an interval of approximately 141 bins, and each bunch is composed of 46 bins. This scheme ensures that about 75% dose reduction is expected and that the practical implementation of it is available.

3) MVUS-Moving

This scheme uses a similar collimator that is used in the MVUS-Bulk, but we reciprocate the collimator during a scan so that the X-rays through each opening in the collimator can sweep the detector bins sinusoidally. The amplitude of the oscillation is designed to cover each partition, i.e. one bunched

detector bins plus the blocked bins corresponding to one interval in the MVUS-Bulk.

We used a total-variation minimization algorithm and micro-CT data of a mouse head for image reconstruction. Each projection data is in a 2-D array of 1120 by 1232 detector pixels, and we used only the mid-rows of the data for image reconstruction of the central slice. The reconstructed image array size is 512 by 512. Various data sampling schemes were numerically realized based on the micro-CT data.

III. RESULTS

Considering the cylindrical symmetry, we analyzed the measures only in the middle row of the image array. Mid-line profiles of the sampling density for all the sampling schemes are shown in Fig. 2. It is noticed that profiles of the sparse-view and the bunched sparse-view are quite uniform compared to others. The MVUS-Fine profile is relatively uniform. The MVUS-Moving line-profile is a bit fluctuating, and the MVUS-Bulk profile shows a gross fluctuation in a quasi-periodic form.

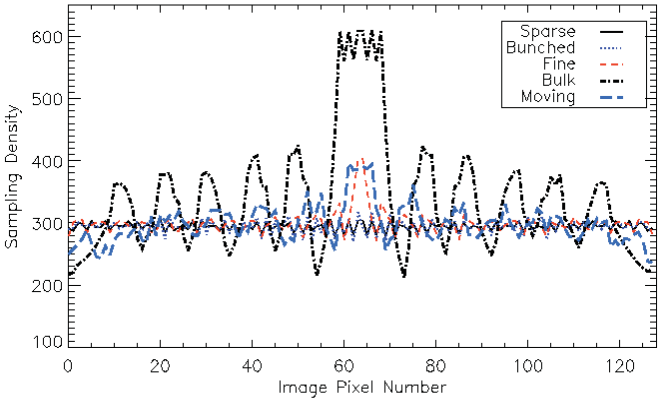


Figure 2. Comparison of mid-line profiles of the sampling density.

Mid-line profiles of the DIC for all the sampling schemes are also shown in Fig. 3. It is observed that DIC values of the sparse-view and the MVUS-Fine are higher than the other cases and that both schemes have pretty close DIC values each other. Profile of the MVUS-Moving shows higher DIC values than those of the MVUS-Bulk and the bunched sparse-view. DIC profile of the MVUS-Bulk shows a similar fluctuating pattern as its sampling density profile shown in Fig. 2. The bunched sparse-view case shows the lowest DIC values out of the investigated schemes.

Reconstructed images of the mouse head from the data prepared by various sampling schemes are shown in Fig. 4. Note that the top portion of each image is trimmed to better utilize the display space in Fig. 4. For comparison, a reference image was also reconstructed by use of the TV minimization algorithm from the full projection data, and is shown in Fig. 4(a). Compared to the reference image, we obtained acceptable quality images in case of the sparse-view, the MVUS-fine, and the MVUS-Moving as shown in Fig. 4(b), 4(d), and 4(f), respectively. Although the image quality assessment should be performed in an objective manner such as detectability measure depending on an imaging task, we provide only a visual comparison here. The reconstructed image from the data acquired with the bunched sparse-view scheme shows pronounced image artifacts as shown in Fig. 4(c). The reconstructed image from the data acquired with the MVUS-Bulk scheme shows ring-shape artifacts as shown in Fig. 4(e).

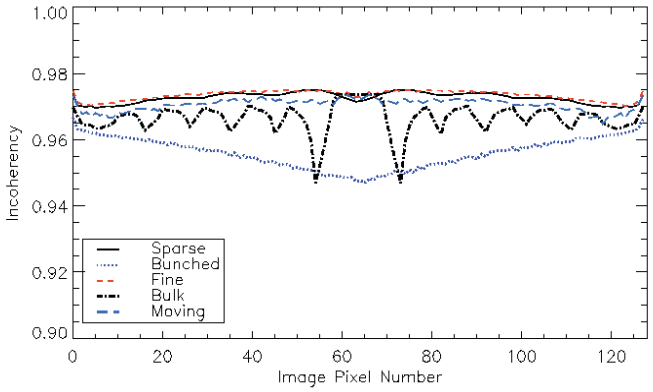


Figure 3. Comparison of mid-line profiles of DIC.

IV. DISCUSION

Images from the data acquired by the sparse-view, MVUS-Fine, and MVUS-Moving sampling schemes were successfully reconstructed, and they are all comparable to the reference image. Interestingly, the DIC values of these three sampling schemes are all high and relatively uniform, but the sampling density varies to a certain degree among them.

In the bunched sparse-view case, the sampling density is quite uniform but the DIC values are low particularly toward the center. Having seen the poor image quality of the bunched sparse-view case after reconstruction, one can conjecture that the DIC plays a dominant role in the image reconstruction in comparison with the sampling density in this work.

The DIC values of the MVUS-Bulk scheme are fluctuating and are relatively higher than those of the bunched sparse-view scheme. However, the reconstructed image quality of the MVUS-Bulk is much worse than that of the bunched sparse-view. We believe that the sampling density nonuniformity of the MVUS-Bulk aggravated the image quality particularly in the regions where poor sampling occurs. Interestingly, the regions that have low sampling density are also subject to low DIC in the MVUS-Bulk case. Additionally, the nonuniformity of sampling density is thought to cause the ring-shape artifacts whereas the low DIC is more related to the streaky artifacts.

Based on these findings, we suggest to take care of the two measures, i.e. sampling density and DIC, when designing a sparse-sampling scheme for low-dose CT. Under a given constraint of dose reduction in the CS-based image reconstruction frame, one should pursue a sparse-sampling scheme that has high DIC and uniform sampling density.

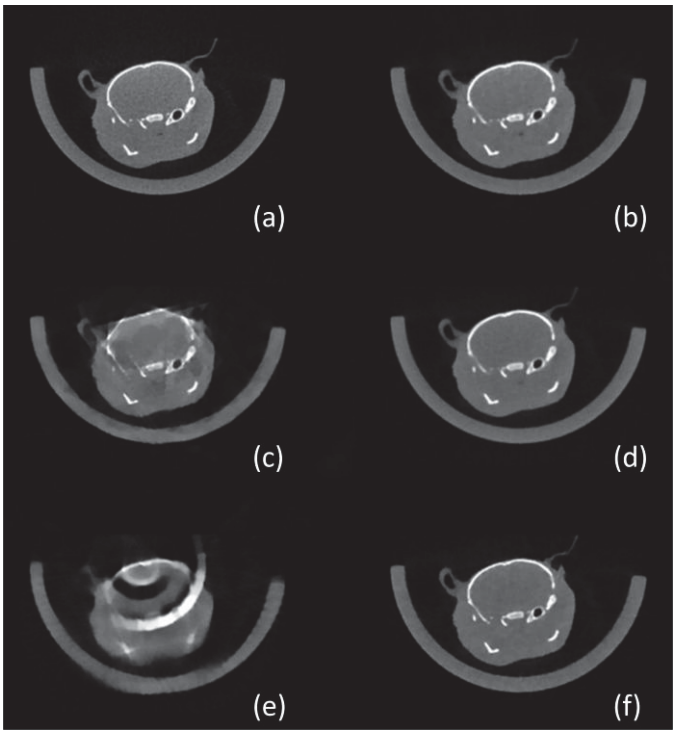


Figure 4. Reconstructed images of a mouse head with various sampling schemes are shown: (a) Reference image, (b) spare-view, (c) bunched sparse-view, (d) MVUS-Fine, (e) MVUS-Bulk, and (f) MVUS-Moving.

For example, when a scanning protocol is designed in the low-dose CT context with the CS-based algorithms, it should be kept in mind that data redundancy is undesirable because the redundancy of data would decrease the data incoherency. Although the MVUS-Bulk scheme assumed regularly partitioned openings with a given size in this work, it has a room to explore for designing the optimum sampling scheme and is actually in progress in our lab. Because of its simplicity in technology, we believe that the MVUS schemes can provide very attractive low-dose scanning options not only to the new CT systems but also to the existing CT systems.

V. CONCLUSION

In this work, we investigated various sparse sampling schemes and their effects on image quality. We proposed to use two measures, i.e. sampling density and data incoherency to assess the utility of each sampling scheme. It is found that both the sampling density and the data incoherency affect the image quality in the CS-based reconstruction. Among the sampling schemes that we investigated, the sparse-view, the MVUS-Fine, and the MVUS-Moving have shown promising results. In various medical imaging applications including diagnostics and interventions, designing a sampling scheme would be very important when a sparse sampling option is chosen for dose reduction. With this strategic approach, one can acquire optimally-sampled sparse data so that the CS-based algorithms can best perform in terms of image quality.

REFERENCES

[1] D. J. Brenner and E. J. Hal, "Computed tomography – an increasing source of radiation exposure", N. Eng J. Med, vol. 357, pp. 2277–2284, 2007.

[2] L. Yu, X. Liu, S. Leng, J. M. Kofler, J.C. Ramirez-Giraldo, M. Qu, J. Christner, J. G. Fletcher and C. H. McCollough, " Radiation dose reduction in computed tomography: Techniques and future perspective", Imaging in Medicine vol. 22, no. 1, pp. 65–84, 2009.

[3] M. J. Slegel, B. Schmldt, D. Bradley, C. Suess, and C. Hlldebolt, "Radiation dose and image quality in pediatric CT: Effect of technical factors and phantom size and shape", Radiology, vol. 233, pp. 512-522, 2004.

[4] D. L. Donoho, "Compressed sensing", IEEE Trans. Info. Theory", vol. 52, pp. 1289–1306, 2006.

[5] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT", J. X-ray Sci. Tech., vol. 14, no. 2, pp. 119–139, 2006.

[6] J. H. Jorgensen, E. Y. Sidky, and X. Pan, " Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT", Med. Phys., arXiv:1109.0629v3, 2012.

[7] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method", Med. Phys., vol. 37, pp. 37:5113–25, 2010.

[8] T. Niu and L. Zhu, "Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: phantom studies", Med. Phys., vol. 39, pp. 4588–98, 2012.

[9] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan, "Optimization-based reconstruction of sparse images from few-view projections", Phys. Med. Bio., vol. 57, pp. 5245–73, 2012.

[10] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT image from highly undersampled projection data sets," Med. Phys., vol. 35, no. 2, pp. 660–663, 2008.

[11] S. Abbas, J. Min, and S. Cho., "Super-sparsely view-sampled cone-beam CT by incorporating prior data", J. X-ray Sci. Tech, 2012 (in press).

[12] S. Cho, T. Lee, J. Min, and H. Chung, "Feasibility study on many-view under sampling (MVUS) technique for low-dose computed tomography", Opt. Eng. vol. 51, no. 8, pp. 080501, 2012.

[13] R. L. Siddon, "Fast calculation of the exact radiological path for a three dimensional CT array", Med. Phys., vol. 12, pp. 252-255, 1985.

Optimized MAP Reconstruction of H^2-weighted Fourier Rebinned TOF PET

Yanguang Lin[1], Bing Bai[2], Wentao Zhu[1], Ran Ren[1], Quanzheng Li[3], Magnus Dahlbom[4], Frank DiFilippo[5], Richard M. Leahy[1]

***Abstract*—Time-of-flight (TOF) information improves signal to noise ratio in Positron Emission Tomography (PET) at higher computation cost. We have previously developed approximate Fourier methods that rebin TOF data into either 3D nonTOF or 2D nonTOF formats. We refer to these methods respectively as FORET-3D and FORET-2D. Here we describe maximum a posteriori (MAP) estimator for use with FORET rebinned data. We first derive approximate expressions for the variance of the rebinned data. We then use these results to rescale the data so that the variance and means are approximately equal allowing us to use the Poisson likelihood model for MAP reconstruction. Using these methods we compare performance of FORET-2D and 3D with TOF and nonTOF reconstructions using phantom and clinical data. Our phantom results show a small loss in contrast recovery at matched noise levels using FORET compared to reconstruction from the original TOF data. Clinical examples show FORET images that are qualitatively similar to those obtained from the original TOF PET data but with a small increase in variance at matched resolution.**

I. Introduction

Over the last two decades, time-of-flight (TOF) information [1] have been used to improve Positron Emission Tomography (PET) image quality. In TOF PET, the difference between the arrival time of the photons is used to better localize the emisssion source along the line of response (LOR). It has been shown that using TOF information, the signal-to-noise ratio (SNR) of the image can be substantially improved [2].

Depending on the timing resolution, using TOF information typically increases the size of the dataset by a factor of 10 to 20. The increased data size significantly increases the PET reconstruction time and presents challenges in fully using TOF data in clinical PET studies.

One way to reduce computational cost is to exploit the redundancy in the data. Generalized from Fourier rebinning methods [3] that rebin 3D nonTOF PET to 2D nonTOF PET data, we have derived a pair of Fourier rebinning methods that rebin TOF singrams to either 2D or 3D nonTOF sinograms [4], which we call FORET-2D or FORET-3D respectively.

[1]Y. Lin, W. Zhu, R. Ran and R. Leahy are with the Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089, USA, email: leahy@sipi.usc.edu

[2]B. Bai is with the Molecular Imaging Center, University of Southern California, Los Angeles, CA 90033, USA

[3]Q. Li is with the Department of Radiology, Harvard Medical School/MGH, Boston, MA 02114, USA

[4]M. Dahlbom is with the Department of Molecular and Medical Pharmacology, University of California, Los Angeles, CA 90095, U.S.A

[5]F. DiFilippo is with the Department of Nuclear Medicine, Cleveland Clinic Main Campus, Cleveland, OH 44195, U.S.A

The FORET rebinned sinogram has a complicated noise structure as opposed to the Poisson noise model for TOF PET data. The reconstruction algorithm needs to consider this noise structure to optimally utilize the SNR gain from TOF information. To achieve this goal, inspired by [5], we first develop an approximate variance model for FORET rebinned data. Using this variance model, which accounts for the effects of random and scatter correction, rebinning and arc correction, we then apply MAP [6] reconstruction methods to the rebinned data. Similar approaches have been applied in FORE reconstruction [7], [8].

Rebinned sinograms also suffer from propogated blurring caused by the detector response as well as additional blurring caused by interpolation, arc correction and other approximations. Resolution in the reconstructed images will be compromised if this blurring effect is not taken into account when reconstructing from rebinned data. Thome and Qi describe a method for estimating the Fourier rebinning blur kernel using point source data [9]. Here, we extend this approach to rebinning of TOF data.

We evaluate the quality of FORET+MAP images using phantom and patient scans. First resolution calibration curves are calculated for each of the reconstruction methods. We then compute MAP reconstructions at matched resolution from FORET rebinned sinograms using the derived noise model and estimated blur kernels. Finally we compare image quality in terms of contrast recovery and variance in a phantom study and conclude with a clinical example.

II. Optimized MAP Reconstruction for FORET Rebinned TOF Data

A. 3D TOF PET Data

FORET rebinning is based on a line integral model and the discretized form assumes uniform sampling with respect to the coordinates of the 3D TOF data. The measured TOF PET data $\mathbf{y}^{TOF}$ can be modeled as independent Poisson random variables with mean and variance given by:

$$\bar{\mathbf{y}}^{TOF} = var(\mathbf{y}^{TOF}) = \mathbf{P}^{TOF}\mathbf{x} + \bar{\mathbf{r}}^{TOF} + \bar{\mathbf{s}}^{TOF} \quad (1)$$

where $\mathbf{x}$ is the 3D tracer distribution or image, and $\bar{\mathbf{y}}^{TOF}$, $\bar{\mathbf{r}}^{TOF}$ and $\bar{\mathbf{s}}^{TOF}$ are the means of the prompt, random and scattered events in the 3D TOF data respectively; $var(\mathbf{y}^{TOF})$ is a vector representing the variance for each LOR. $\mathbf{P}^{TOF}$ is the system matrix which can be written as:

$$\mathbf{P}^{TOF} = \mathbf{P}^{TOF}_{norm}\mathbf{P}^{TOF}_{blur}\mathbf{P}^{TOF}_{attn}\mathbf{P}^{TOF}_{geom} \quad (2)$$

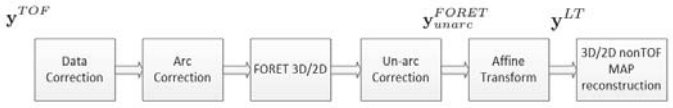


Fig. 1. Flow chart for MAP reconstruction using FORET rebinned data.

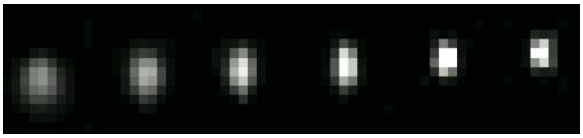


Fig. 2. Blur kernel used for FORET rebinned data, from left to right: larger radial displacement to smaller radial displacement.

where $\mathbf{P}^{TOF}_{geom}$ is the geometric projection matrix for TOF data, which is based on solid angle calculations with sensitivities for each voxel weighted by the TOF kernel [6], [10]. $\mathbf{P}^{TOF}_{norm}$ and $\mathbf{P}^{TOF}_{attn}$ are the diagonal normalization and attenuation matrices. $\mathbf{P}^{TOF}_{blur}$ operates on the data in sinogram space to model detector response as a local blurring.

B. Noise Properties of FORET Rebinned Data

To account for the discrepancies between the line integral and accurate system models above, we need to perform a sequence of operations on the data as illustrated in Figure 1, first to convert to a form suitable for FORET rebinning, and then back to a form that can be combined with MAP reconstruction of the 2D or 3D nonTOF data.

The FORET rebinned data after unarc correction in Figure 1 can be written as

$$\begin{aligned}&\mathbf{y}^{FORET}_{unarc} \\ =&\mathbf{M}_{unarc}\mathbf{R}\mathbf{M}_{arc}\mathbf{P}^{TOF\ -1}_{norm}\mathbf{P}^{TOF-1}_{attn}(\mathbf{y}^{TOF}-\bar{\mathbf{r}}^{TOF}-\bar{\mathbf{s}}^{TOF})\end{aligned} \tag{3}$$

where $\mathbf{y}^{FORET}_{unarc}$ represents the unarc corrected 2D or 3D nonTOF data. $\mathbf{M}_{arc}$ and $\mathbf{M}_{unarc}$ are two matrices representing the arc and unarc corrections and $\mathbf{R}$ is the matrix representing FORET rebinning; $\bar{\mathbf{r}}^{TOF}$ and $\bar{\mathbf{s}}^{TOF}$ are the estimated mean random and scatter sinograms.

The mean of the rebinned data is

$$\bar{\mathbf{y}}^{FORET}_{unarc} \approx \mathbf{M}_{unarc}\mathbf{R}\mathbf{M}_{arc}\mathbf{P}^{TOF}_{blur}\mathbf{P}^{TOF}_{geom}\mathbf{x} \tag{4}$$

Although FORET is based on the line integral geometry model, blurring associated with the detector response remains in the data after rebinning. We therefore model the rebinned data as:

$$\bar{\mathbf{y}}^{FORET}_{unarc} \approx \mathbf{P}_{blur}\mathbf{P}_{geom}\mathbf{x} \tag{5}$$

where $\mathbf{P}_{blur}$ is the estimated FORET blurring kernel (Fig. 2) which contains the propogated detector response blurring kernels and the additional blurring introduced by FORET rebinning. It was estimated from simulated point source data using an expectation maximization (EM) algorithm [9]. $\mathbf{P}_{geom}$ represents the 3D or 2D geometric projection matrix for FORET-3D and FORET-2D respectively.

Comtat et al estimated the variance of FORE rebinned nonTOF data [5]. Similarly we can derive expressions for variance of the FORET rebinned TOF data as:

$$var(\mathbf{y}^{FORET}_{unarc}) = \mathbf{G}\bar{\mathbf{y}}^{FORET}_{unarc} + \mathbf{L}(\bar{\mathbf{r}}^{3D} + \bar{\mathbf{s}}^{3D}) \tag{6}$$

where $\bar{\mathbf{r}}^{3D}$ and $\bar{\mathbf{s}}^{3D}$ are respectively the 3D estimates of randoms and scatters. $\mathbf{G}$ and $\mathbf{L}$ are respectively matrices combining all terms that result from the effect of applying the operator defined in (3) to the corrected data, which results in the propagation of variance in the trues, scatters and randoms into the final rebinned variance.

Note that in equation (6) the rebinned sinogram are no longer Poisson distributed. In particular, the variance of the sinogram is typically significantly smaller than its mean. A direct MAP reconstruction from the rebinned sinogram will yield a suboptimal solution because of the inaccurate noise model. To compensate for the SNR improvement, we apply an affine transform to the rebinned sinogram so that the transformed data are approximately Poisson and then perform MAP reconstruction on these transformed data.

The affine transformed data can be written as:

$$\mathbf{y}^{LT} = \mathbf{D}\mathbf{y}^{FORET}_{unarc} + \mathbf{d} \tag{7}$$

where $\mathbf{D}$ is a diagonal matrix and $\mathbf{d}$ a vector. It is straightforward to show that the mean and variance of the transformed data are equal when

$$\begin{aligned}\mathbf{D} &= \mathbf{G}^{-1} \\ \mathbf{d} &= \mathbf{G}^{-2}\mathbf{L}(\bar{\mathbf{r}}^{3D} + \bar{\mathbf{s}}^{3D})\end{aligned} \tag{8}$$

This affine transformation is applied to the rebinned data prior to MAP reconstruction. The Poisson likelihood function used in the MAP optimization is then based on the model for the mean of the data:

$$\bar{\mathbf{y}}^{LT} = \mathbf{D}\bar{\mathbf{y}}^{FORET}_{unarc} + \mathbf{d} = \mathbf{D}\mathbf{P}_{blur}\mathbf{P}_{geom}\mathbf{x} + \mathbf{d} \tag{9}$$

III. Results

A. Variance Estimation for FORET Rebinned Data

In this section, we use a realistic Monte Carlo simulation to verify the variance calculation of FORET rebinned sinograms (6). We simulated the geometry of the Siemens mCT TOF PET scanner [11]. We selected the image for one bed position extracted from a whole-body patient study and smoothed with a 5mm Gaussian filter. We then forward projected the smoothed image using the full system model to create noiseless 3D TOF sinograms [10]. The normalization, attenuation correction, randoms and scatters calculated from the scanner software were used in the simulation. We then generated 100 datasets from this mean data set using a Poisson random number generator and rebinned them using our FORET methods.

Figure 3 shows profiles of the sample variance from the Monte Carlo simulation and theoretical computation from (6) for FORET-3D and FORET-2D. While there is still residual Monte Carlo variability in the variance estimates, it is clear that for both 2D and 3D rebinning, the variance estimates are approximately correct. To look at these relationships in a more general perspective, Figure 4 shows the histograms of the ratio between mean and variance of the data after the affine transform using (7) at which point the mean and variances should be equal. In both cases the ratios have their mode

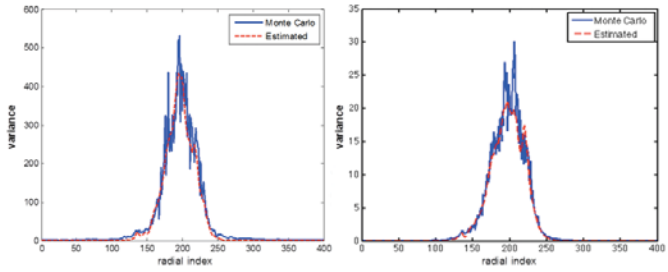


Fig. 3. Comparison of sample variance from Monte Carlo simulation and estimated from (6), angle index=100, sinogram index=5. Left: FORET-3D, Right: FORET-2D.

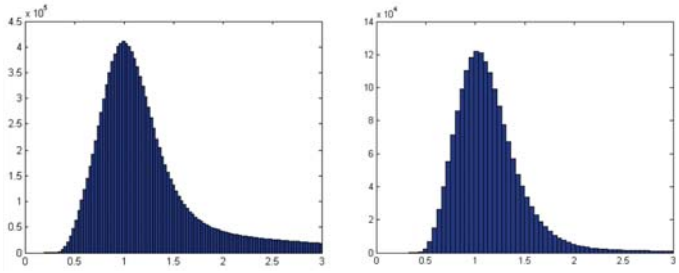


Fig. 4. Histogram of mean to variance ratios of FORET after the affine transform in (7) at which point the mean and variance should be equal. Left:FORET-3D, Right: FORET-2D.

close to unity. The variability about this mode is in part due to systematic errors in the variance approximation but also residual variance in the variance estimated using the Monte Carlo simulation (as also seen in Figure 3).

B. Resolution vs. β Calibration

We simulated a line source in a uniform cylindrical phantom with low background activity concentration(ratio 300:1) and reconstructed with different values of β to compute the resolution vs. β calibration curves [12]. Results are shown in Figure 5. These curves clearly show that dependence on β varies with processing method. Note that as we reduce dimensionality from 3D TOF to 3D nonTOF (either through FORET-3D or summing over TOF bins) to 2D nonTOF (through FORET-2D), larger values of β are required to meet a given resolution for the lower resolution portion of the curves. At higher resolutions the 3D and 2D nonTOF curves cross. At these higher resolutions it is possible that the different approximations as a function of rebinning scheme have a larger impact on resolution resulting in the cross over of the curves.

C. Phantom Measurements

Reconstructed image quality was evaluated using an anthropomorphic torso phantom with breast attachments (Data Spectrum, Hillsborough, NC). The phantom was filled with ^{18}F-FDG and scanned for 10 minutes on a Siemens mCT scanner. The phantom has liver and myocardium compartments and two lesions represented using hot spheres with 15.9 mm and 12.7 mm diameter inserted in the abdominal region. The relative activity concentration in the torso background, liver, myocardium and lesions was 1:1.91:3.85:3.85.

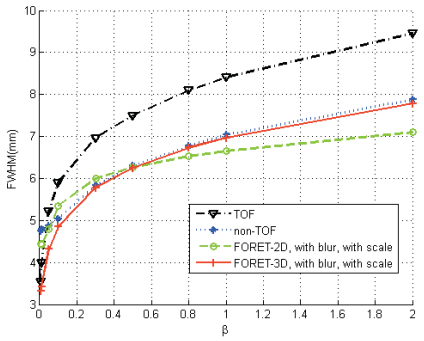


Fig. 5. Resolution vs. β calibration table generated from simulated line source phantom data for the Siemens mCT scanner for each of the four different data representations used in subsequent studies. voxel size is 2mm.

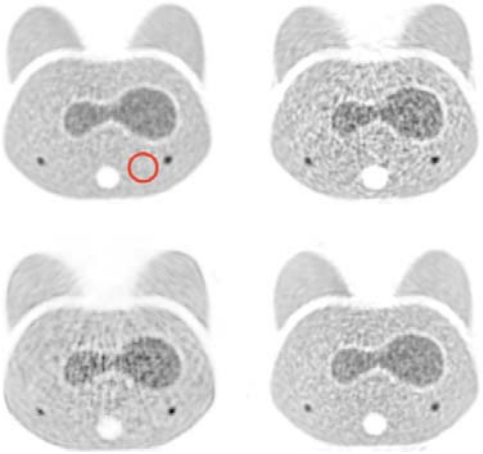


Fig. 6. Sample reconstructed phantom images using different data formats. Upper left: TOF, upper right: nonTOF, lower left: FORET-2D, lower right: FORET-3D.

Sample transaxial images reconstructed from 3D TOF data, 3D nonTOF, FORET-3D and FORET-2D data are shown in Figure 6. An 8mm diameter region of interest (ROI) was drawn in the center of the hot sphere on the right and a 24mm diameter spherical ROI drawn in the body (red circle). We measured the contrast recovery coefficient (CRC) using the equation $CRC = (\bar{f}_{hot} - \bar{f}_{bkg})/(a-1)$, where $\bar{f}_{hot}$ and $\bar{f}_{bkg}$ are the average values of the voxels in the hot ROI and background ROI respectively; a is the true activity ratio ($a = 3.85$). The CRC vs. noise result is shown in Figure 7. It can be seen that the TOF images have the best quality, with lower noise at the same contrast recovery level, while FORET-3D and FORET-2D result in some loss of contrast at matched noise levels, but still noticeably better than the nonTOF reconstructions. The FORET-2D image also shows artifacts which may arise from the less accurate approximation used in FORET-2D than in FORET-3D. We also reconstructed the FORET data without the affine transform for variance matching and without blur kernels. This clearly shows worse performance than all other methods and indicates the importance of response modeling and variance scaling.

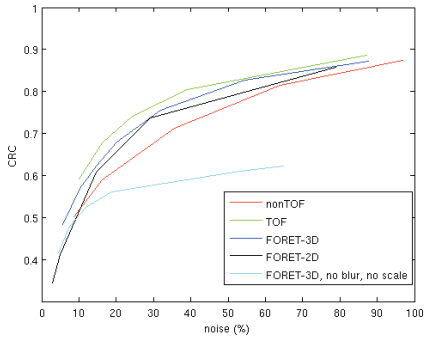


Fig. 7. Noise vs. contrast recovery curves of images reconstructed from TOF, nonTOF, FORET-3D and FORET-2D data.

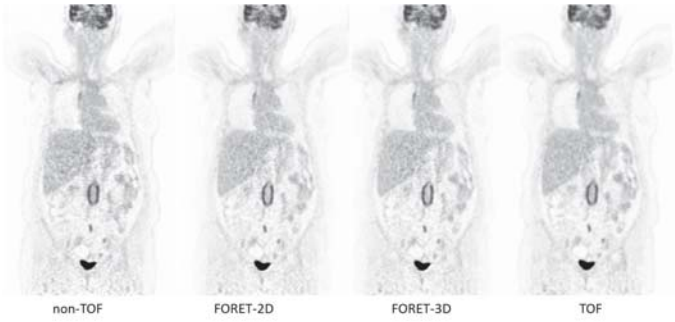


Fig. 8. Comparison of whole-body patient images. All images have resolution of 6mm.

D. Whole-body Patient Scan

To illustrate the relative performance of the rebinning methods in clinical studies we applied all four reconstruction methods to a clinical whole-body scan collected using the Siemens mCT scanner. The patient was a 66-year-old female, weighting 85.7kg. 451 MBq of ^{18}F-FDG was injected and the scan started 46 minutes after injection. The patient was scanned with six bed positions and three minutes per bed. The smoothing parameter β was selected such that the resolution of the reconstructed image was 6mm. In all the reconstructions, we ran 1 iteration of 3D OSEM, which was used as initial image for MAP reconstruction. We then ran 20 iterations of MAP to ensure effective convergence. Coronal sections through these images are shown in Figure 8. The ordering of image quality is consistent with the phantom results.

IV. Conclusion

We have optimized MAP reconstruction for Fourier rebinned TOF PET data. We derived a noise model for the FORET rebinned data, and used this as the basis for an affine transform such that the transformed data are approximately Poisson. Monte Carlo simulation results show that the variance calculations, although with several approximations, are reasonably accurate. Phantom studies confirmed that using the new noise model can improve image quality.

Our results show a clear progression in quality from TOF through FORET-3D, FORET-2D and finally nonTOF. This is clear in both the phantom and clinical examples. While the ordering is unsurprising, we expected that the FORET results, particularly for 3D, would be closer in quality to TOF than they are. Our basis for this assumption is that our earlier results in [13] indicated that there was little information loss in FORET rebinning. The cause of the somewhat worse than expected performance for FORET may be due to correlation introduced into the rebinned data which we do not account for in the scaling process described here. The incorporation of correlation information in nonTOF rebinning has been previously studied [7]. We plan to explore similar approaches to improve our noise model for FORET rebinned data in future. Despite this issue, however, the FORET rebinned images still retain much of the advantage of TOF data in comparison to nonTOF data.

Acknowledgements

The authors thank Michael E. Casey, Vladimir Panin and Christian Michel of Siemens Medical Solutions for their assistance in working with data from Siemens mCT scanner. This work was supported by NIH grant no. R01 EB010197.

References

[1] M. Conti, B. Bendriem, M. Casey, M. Chen, F. Kehren, C. Michel, and V. Panin, "First experimental results of time-of-flight reconstruction on an lso pet scanner," *Physics in Medicine and Biology*, vol. 50, no. 19, p. 4507, 2005.

[2] K. Vunckx, L. Zhou, S. Matej, M. Defrise, and J. Nuyts, "Fisher information-based evaluation of image quality for time-of-flight pet," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 2, pp. 311–321, 2010.

[3] M. Defrise, P. Kinahan, D. W. Townsend, C. Michel, M. Sibomana, and D. Newport, "Exact and approximate rebinning algorithms for 3-d pet data," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 2, pp. 145–158, 1997.

[4] S. Cho, S. Ahn, Q. Li, and R. M. Leahy, "Exact and approximate fourier rebinning of pet data from time-of-flight to non time-of-flight," *Physics in Medicine and Biology*, vol. 54, no. 3, p. 467, 2009.

[5] C. Comtat, P. Kinahan, M. Defrise, C. Michel, and D. Townsend, "Fast reconstruction of 3d pet data with accurate statistical modeling," *Nuclear Science, IEEE Transactions on*, vol. 45, pp. 1083 –1089, jun 1998.

[6] J. Qi, R. Leahy, C. Hsu, T. Farquhar, and S. Cherry, "Fully 3d bayesian image reconstruction for the ecat exact hr+," *Nuclear Science, IEEE Transactions on*, vol. 45, pp. 1096 –1103, jun 1998.

[7] A. Alessio, K. Sauer, and C. A. Bouman, "Map reconstruction from spatially correlated pet data," *Nuclear Science, IEEE Transactions on*, vol. 50, no. 5, pp. 1445–1451, 2003.

[8] X. Liu, C. Comtat, C. Michel, P. Kinahan, M. Defrise, and D. Townsend, "Comparison of 3-d reconstruction with 3d-osem and with fore+ osem for pet," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 8, pp. 804–814, 2001.

[9] M. S. Tohme and J. Qi, "Iterative reconstruction of fourier-rebinned pet data using sinogram blurring function estimated from point source scans," *Medical physics*, vol. 37, p. 5530, 2010.

[10] Y. Lin, Q. Li, and R. M. Leahy, "Fast gpu-based time-of-flight map reconstruction with a factored system matrix," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pp. 2889–2893, IEEE, 2010.

[11] B. W. Jakoby, Y. Bercier, M. Conti, M. E. Casey, B. Bendriem, and D. W. Townsend, "Physical and clinical performance of the mct time-of-flight pet/ct scanner," *Physics in Medicine and Biology*, vol. 56, no. 8, p. 2375, 2011.

[12] Q. Li, B. Bai, S. Cho, A. Smith, and R. Leahy, "Count independent resolution and its calibraion," in *Proceedings of 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp. 223–226, 2009.

[13] S. Ahn, S. Cho, Q. Li, Y. Lin, and R. Leahy, "Optimal rebinning of time-of-flight pet data," *Medical Imaging, IEEE Transactions on*, vol. 30, pp. 1808 –1818, oct. 2011.

The Effect of Motion on Joint Estimates of Activity and Attenuation from Time-of-Flight PET Data

Ahmadreza Rezaei[1], Johan Nuyts[1], and Michel Defrise[2]

Abstract— **Recent studies show that joint reconstruction of activity and attenuation is possible with time-of-flight PET data. However, when there is motion during the acquisition of the emission data, the properties of the reconstructions are not known. In this study, we classify three theoretical types of motion and analyze joint reconstructions when the emission data has been affected by each type separately. We use the existing TOF-MLAA algorithm for this purpose and observed that motion during the scan can make the data inconsistent, such that the resulting reconstructions are not only affected by motion blur, but also suffer from artifacts.**

I. Introduction

It is known that motion in PET and PET/CT can result in errors due to incorrect attenuation correction in addition to motion blur in the estimated emission image [1]. In [2], it was shown that a difference in emission and transmission resolution can also introduce artifacts due to attenuation correction of the emission data. Different methods have been proposed to correct the emission data affected by motion during the scan when gated data are not available [3]–[5]. In gated PET studies however, it is preferable to acquire a single CT scan, register the attenuation images non-rigidly to the PET images before attenuation correction and correct for attenuation in the data using the deformed (registered) CT scans for each phase in the cardiac/respiratory cycle [6]–[8]. Recent studies have shown that when time-of-flight (TOF) data are available, joint reconstruction of the activity and attenuation is possible up to a scale factor [9]–[11]. Hamill and Panin [5] have shown that when motion occurs during the PET-scan, MLAA produces a blurred mu-map, yielding better attenuation correction than is obtained from a fast, motion-free CT scan. However, the exact effect of motion during acquisition on the joint reconstruction of activity and attenuation is not clearly known. In order to further investigate the effects of motion, we identify different types of motion that theoretically could occur during the emission measurements, simulate each type and analyze the final activity and attenuation reconstructions. We use the existing algorithm for maximum likelihood estimation of activity and attenuation (MLAA) [10] and compare the reconstructed images to reference ML images. The reference activity image is produced with MLEM (maximum likelihood expectation maximization) assuming that the attenuation is known, and the reference attenuation image is computed with MLTR (maximum likelihood for transmission tomography) assuming knowledge about the true tracer distribution. Our main result is that TOF-MLAA produces consistent pairs of activity and attenuation reconstructions that best describe the emission data when the motion either affects the activity or the attenuation (first 2 motion types), but not both. Our simulations indicate that in the case of simultaneous motion in activity and attenuation (type 3 motion), the emission data are no longer consistent, i.e. there exists no value for the estimated parameters (here the activity and the attenuation) that would produce the measured data, using the current acquisition model (the model here assumes a motion free object). Thus, TOF-MLAA can only reconstruct the consistent portion of the emission data. Nevertheless, we see that the region which is not affected by the motion can be quantitatively reconstructed. In the case of inconsistencies in the emission sinogram, it is not known whether the emission measurements could be split into multiple consistent emission measurements or if it is possible to retrieve information about the motion by analyzing the emission measurements.

[1] Dept. of Nuclear Medicine, Katholieke Universiteit Leuven, B-3000, Leuven, Belgium, [2]Dept. of Nuclear Medicine, Vrije Universiteit Brussel, B-1090 Brussels, Belgium. E-mails: ahmadreza.rezaei@uz.kuleuven.be, johan.nuyts@uz.kuleuven.be, and mdefrise@vub.ac.be

[3]This research is supported by a research grant (GOA) from K.U.Leuven.

II. Methods

In order to have a better understanding of the effect of motion in the joint estimation problem, we analyze three types of motion, i.e. 1- motion in the activity image, 2- motion in the attenuation image and 3- simultaneous motion in both the activity and attenuation images. Although the first two types of motion may not be as common as the third type (type 2 motion in particular seems rather unrealistic), the analysis gives additional insight into the problem of motion in joint image reconstruction. We will simulate the motion in successive frames, make projections of each frame and sum over all projections. We then attempt to reconstruct the motion affected measurements and compare the results with the total activity (summed over all frames) and average attenuation (average over all frames).

We model the TOF-PET expected counts $\bar{y}_{it} = a_i p_{it}$ for line of response (LOR) i and TOF-bin t as the multiplication of the (unattenuated) TOF projections of the activity image p_{it} and the attenuation factors a_i which are computed as

$$p_{it} = \sum_{j=1}^{J} c_{ijt}\lambda_j, \quad a_i = e^{-\sum_{j=1}^{J} l_{ij}\mu_j} \tag{1}$$

where λ_j and μ_j are the activity and attenuation coefficient at voxel j, J is the total number of voxels, c_{ijt} is the sensitivity of the measurement bin at (i,t) for activity in j in absence of attenuation and l_{ij} is the intersection length of LOR i with voxel j.

A. Activity Motion

The measurements due to this type of motion can be expressed as,

$$\bar{y}_{it} = \sum_{f=1}^{F} \bar{y}_{it}^{f} = a_i \sum_{f=1}^{F} p_{it}^{f} \tag{2}$$

where the superscript f determines the time frame and F is the total number of time frames.

In this type of motion, the attenuation image remains stationary over all time frames Thus reconstructing the measurements $\bar{y}_{it}$, we expect to obtain the stationary attenuation image together with the total activity image.

B. Attenuation Motion

The measurements due to this type of motion can be expressed as,

$$\bar{y}_{it} = \sum_{f=1}^{F} \bar{y}_{it}^{f} = F p_{it} \sum_{f=1}^{F} \frac{a_i^f}{F} \tag{3}$$

In this type of motion, the activity is stationary. Reconstructing the measurements $\bar{y}_{it}$, we expect to obtain F times the stationary activity image together with some "equivalent" attenuation image. Some attenuation correction methods use the average of CT scans extrapolated to the required photon energy of 511 keV. However, it can be seen from eq (3) that the time averaging happens over the attenuation factors, and not over the attenuation images [6].

C. Activity and Attenuation Motion

In this type of motion, since neither the activity image nor the attenuation images are stationary, the measurements $\bar{y}_{it}$ can not be expressed in any simpler form. We define the "residual error" (TOF) sinogram as the difference between the measured data y_{it} and the expected data $\bar{y}_{it}$ corresponding to the forward projection of the activity and attenuation reconstructions. These residuals were very different from zero, indicating that the measurements y_{it} affected by this motion are no longer consistent under a motion-free reconstruction model.

III. Experimental Design and Results

Just to keep things simple, we simulated the activity and attenuation images in three different time frames (F=3), figure 1. A disk-like object was moved within the object to simulate the three types of motions discussed. The disk-like object was set to the background value to simulate the stationary attenuation and activity phantoms in the first two types of motion.

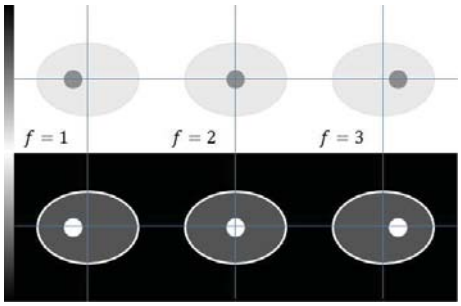


Fig. 1. Activity (top) and attenuation (bottom) images at three different frames (left, center and right) used in the study.

For the simulations below, 2D TOF emission data were generated and organized in 200 radial bins, 168 azimuthal angles, 13 TOF-bins of 312 ps width and with an effective TOF resolution of 580 ps. An oversampling of 3 (i.e. 3 rays per LOR, 9 sub-pixels per image pixel) was used during the simulation, to avoid an exact match between the simulation and the (back)projection during reconstruction. The activity and attenuation images were then reconstructed in a 200×200 pixel grid with a pixel size of 4.1×4.1 mm^2. Activity and attenuation reconstructed using MLAA are then compared to activity reconstructions of MLEM with known attenuation and to attenuation reconstructions of MLTR with known activity, respectively. The MLAA reconstructions are after 3 iterations of 28 subsets, cycling over the attenuation image 5 times for each update of the emission image [10]. The reconstructions of MLEM are after 3 iterations of 28 subsets and the reconstructions of MLTR are after 15 iterations of 28 subsets.

A. Activity Motion - Type I

Since the attenuation image is stationary in this type of motion, the exact attenuation image was projected and used for attenuation correction of the reference MLEM emission reconstruction. We also used the sum of all the activity images as a blank scan to reconstruct the reference attenuation image using the MLTR algorithm. Figure 2 shows horizontal profiles through the center of the reconstructed activity and attenuation images from MLTR, MLEM and MLAA.

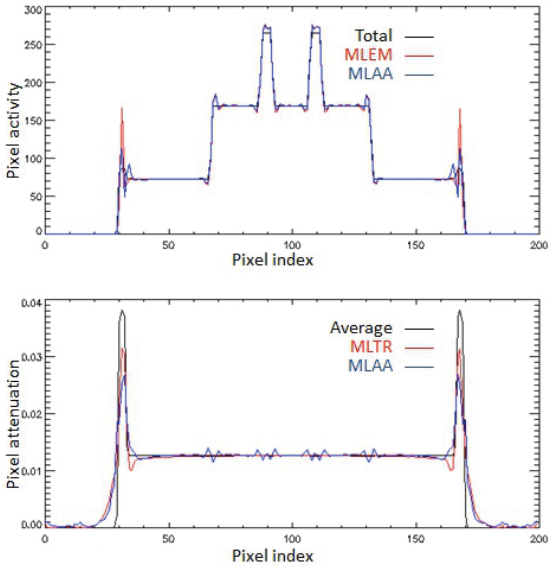


Fig. 2. Horizontal activity (top) and attenuation (bottom) profiles through the MLTR, MLEM and MLAA reconstructions affected by the type I motion. The ML activity and attenuation profiles are compared to the total activity and the average attenuation image of the different time frames, respectively.

The profiles reveal a close agreement between the MLAA reconstructions and the ML reference images. As expected, it can be seen that motion of the activity during the scan results in smoothness (motion blur) of the reconstructed activity image.

B. Attenuation Motion - Type II

In the type II motion, the activity image is stationary. Hence, we use its projections (multiplied by F) as a blank scan to reconstruct the reference attenuation image using MLTR. To simulate the approach where the resolution of the attenuation and activity are matched, we use the projections of the average attenuation image as the corresponding attenuation

that needs to be accounted for during MLEM reconstructions of the activity. Figure 3 shows similar profiles through MLTR, MLEM and MLAA reconstructions of attenuation and activity.

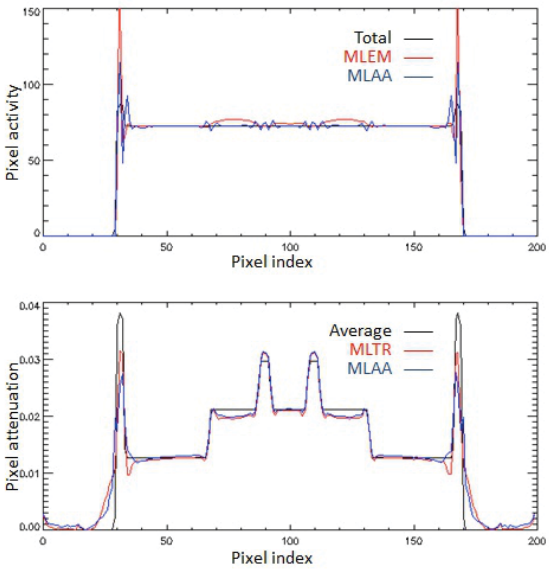


Fig. 3. Horizontal activity (top) and attenuation (bottom) profiles through the MLEM, MLTR and MLAA reconstructions affected by the type II motion.

It can be seen that given the total activity images over multiple frames, MLTR produces an attenuation that is different from the average attenuation. Interestingly, we see that MLAA has been able to produce an attenuation profile similar to the one of MLTR. We will refer to this as an "equivalent" attenuation image, which corresponds to the attenuation image best describing the attenuation correction factors averaged over the time frames as in eq (3) (as opposed to an average over the attenuation images). The difference seen between the average attenuation image and the equivalent attenuation image is due to the nonlinearity of the exponential law of Beer-Lambert.

Figure 3 also shows that the activity profile that was produced using MLEM and the average attenuation image suffers from artifactual increase of the tracer activity in the region which has been affected by this type of motion. In contrast, MLAA was able to produce activity and attenuation images that are better consistent with the emission measurements.

C. Activity and Attenuation Motion - Type III

In this type of motion, neither the activity nor the attenuation images are stationary. However, we will still use the projections of the total activity (which we optimistically would like to recover) to reconstruct an attenuation image with MLTR and use the average attenuation image, as before, to reconstruct the activity image with MLEM. Figure 4 shows the two horizontal profiles of the reconstructed activity and attenuation images from MLTR, MLEM and MLAA.

The activity profiles of figure 4 show that neither MLAA nor MLEM with knowledge of the average attenuation image accurately estimate the total tracer distribution image. Furthermore, MLAA and MLTR with knowledge of the total activity produced different horizontal attenuation profiles. We have also tried reconstructing the total activity using MLEM with the "equivalent" attenuation image, however the result differed

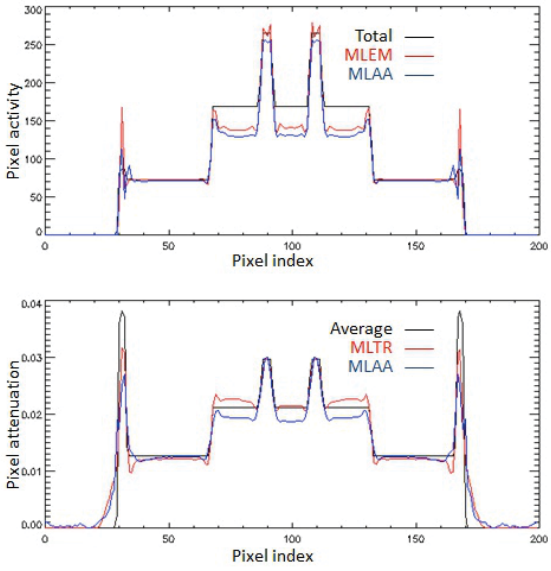


Fig. 4. Horizontal activity (top) and attenuation (bottom) profiles through the MLEM, MLTR and MLAA reconstructions affected by the type III motion.

significantly from the true total activity and was also slightly different from the MLAA activity (profile not shown here).

D. Inconsistencies Due to Motion

The residual error sinograms were calculated for the three motion cases, and we observed signs of inconsistencies in the case of the type III motion. Figure 5 shows (non-TOF) sinograms of the mean and variance, computed by averaging the residual error sinograms over the 13 TOF-bins.

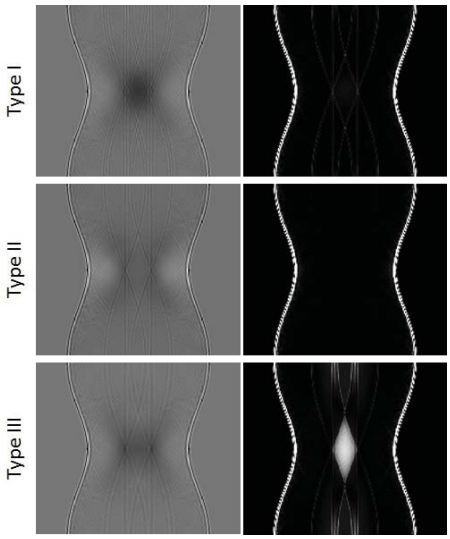


Fig. 5. Mean (left) and variance (right) over the TOF-bins of the residual error sinogram for the type I (top), type II (middle) and type III (bottom) motions. The grey scale is identical for all images.

Figure 5 shows that when there is motion in the activity and attenuation images (type III motion), the mean TOF-bin error does not show significant signs of inconsistencies. However, the variance image shows that in some LORs, there are fairly large positive and negative errors, indicating that for this type of motion the emission data are no longer consistent. We believe that this is why MLAA has not been able to accurately estimate the total activity image in the type III motion.

Fig 5 also shows that the residuals are much larger near the boundary of the sinogram. This is because these LORs have a short intersection with the object, and therefore have few counts, and provide only very limited information. This issue is better described in [9].

E. Locality of the Motion-Affected Region

In the case of the type III motion, we analyzed the image region that is affected. The same phantom of figure 1 was used with the addition of extra details to the phantom. MLAA activity and attenuation reconstructions after 10 iterations and 28 subsets were compared to the total activity and average attenuation, respectively.

Figure 6 shows the reference images, the MLAA reconstructions and the absolute value of the differences between both. The difference images were multiplied with 10 and displayed with the same intensity window as the corresponding reconstructions.

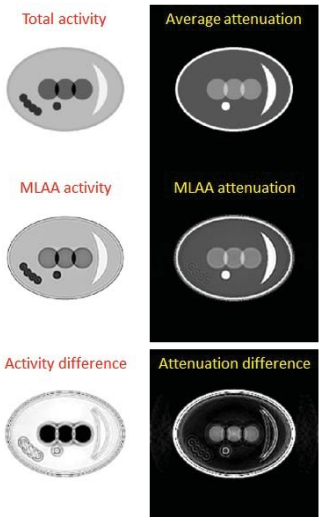


Fig. 6. Reference (top), reconstructed (middle) and ten times the error (bottom) between the activity (left) and attenuation (right) images affected by type III motion. The three images in each column are displayed with the same intensity window.

Figure 6 shows that motion affected the reconstructions locally, while the regions not affected by motion were reconstructed with a good quantitative accuracy.

IV. Discussion and Conclusion

TOF-PET data determine the attenuation sinogram up to a constant [9], and joint estimation of attenuation and activity with a maximum likelihood algorithm was found to be successful and surprisingly stable [10]. In clinical PET/CT imaging, often attenuation correction artifacts are observed, which are due to a geometric mismatch between CT and PET. This mismatch can be caused by patient motion between the CT and PET scans and by motion during the PET scan. The joint estimation of attenuation and emission from the same scan data obviously eliminates between-scan motion. Joint estimation is also assumed to provide an "optimal" reconstruction in the case of in-scan motion, where "optimal" means that the solution maximizes the likelihood. However, it was not clear to what extent this "optimal" activity reconstruction would correspond to the best image that one can hope to obtain from a single ungated scan with motion, which is a motion blurred version of the true activity image.

In this work, we classify motion in the measurements into three types of motion, activity motion, attenuation motion and simultaneous activity and attenuation motion. In the first two types of motion, it was observed that MLAA was able to reconstruct a pair of activity and attenuation images that agreed (almost) exactly with the emission data. The MLAA activity and attenuation images are quantitatively close to MLEM reconstructions of activity with known attenuation and MLTR reconstructions of attenuation with known activity. In the case of type II motion, it was also shown that using the projections of the average attenuation images can result in motion errors in the reconstructed activity image.

In the case of type III motion, the MLAA activity and attenuation profiles were significantly different from the activity reconstruction of MLEM and the attenuation reconstruction of MLTR in the motion affected regions of the reconstructions. Apparently, in this case the emission data are no longer consistent and MLAA has been able to reconstruct only the consistent parts of the emission data. The inconsistency of the data affected by the type III motion is suggested by large values in the TOF residuals sinogram. In our simulations, it appears that LORs which intersect the moving object the most, happen to be the ones that see a high variance in the residual error sinogram. Further investigation is required to check whether or not is it possible to split an inconsistent sinogram into multiple consistent ones or to retrieve information about the motion by analyzing the inconsistencies in the residual errors.

References

[1] M Osman, C Cohade, Y Nakamoto, R Wahl, "Respiratory motion artifacts on PET emission images obtained using CT attenuation correction on PET-CT", *Eur J Nucl Med Mol Imaging*., 2003, 30(4):603-606.

[2] A Chatziioannou, M Dahlborn, "Detailed investigation of transmission and emission data smoothing protocols and their effects on emission images", *IEEE Trans. Nucl. Sci*., 1996, 43(1):290-294.

[3] T Pan, O Mawlawi, S Nehmeh, *et al*., "Attenuation Correction of PET Images with Respiration-Averaged CT Images in PET/CT", *J Nucl Med*, 2005, 46(9):1481-1487.

[4] R Cook, G Carnes, T Lee, G Wells, "Respiration-Averaged CT for Attenuation Correction in Canine Cardiac PET/CT", *J Nucl Med*, 2007, 48(5):811-818.

[5] J Hamill, V Panin, "TOF-MLAA for Attenuation Correction in Thoracic PET/CT", *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim*, 2012, paper M23-1.

[6] J Hamill, G Bosmans, A Dekker, "Respiratory-gated CT as a tool for the simulation of breathing artifacts in PET and PET/CT", *Med Phys*, 2008, 35(2):576-585.

[7] M Dawood, F Buther, X Jiang, K Schafers, "Respiratory Motion Correction in 3-D PET Data With Advanced Optical Flow Algorithms", *IEEE Trans. Med. Imag*., 2008, 27(8):1164-1175.

[8] S McQuade, T Lambrou, B Hutton, "A novel method for incorporating respiratory-matched attenuation correction in the motion correction of cardiac PETCT studies", *Phys. Med. Biol*., 2011, **56**:2903-2915.

[9] M Defrise, A Rezaei, J Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant", *Phys. Med. Biol*., 2012, **57**:885-899.

[10] A Rezaei, M Defrise, G Bal, C Michel, M Conti, C Watson, J Nuyts, "Simultaneous reconstruction of activity and attenuation in Time-of-flight PET", *IEEE Trans. Med. Imag*., 2012, 31(12):2224-2233.

[11] J Nuyts, A Rezaei, M Defrise, "ML-reconstruction for TOF-PET with simultaneous estimation of the attenuation factors", *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim*, 2012, paper M04-1.

Mitigating cone-beam artifacts in short-scan CT imaging for large cone-angle scans

Jed D. Pack[1], Zhye Yin[1], Kai Zeng[2], Brian E. Nett[3]
[1] GE Global Research Center, One Research Circle, Niskayuna, NY USA
[2] Formerly with GE Global Research Center, One Research Circle, Niskayuna, NY USA
[3] GE Healthcare, 3000 N. Grandview Blvd., W-1180, Waukesha WI USA

***Abstract*—It is well established that when reconstructing an image volume from a single circular CT scan (or a portion thereof), cone-beam artifacts result. We describe three categories of cone-beam artifacts. When less than a full scan of data is available (as is typically the case in, e.g., cardiac CT or C-arm volume imaging) and a traditional short-scan FBP-type algorithm is employed, the artifacts from data mishandling usually dominate over the bulk of the reconstructed volume. We present an approach that is fairly straight-forward and provides very good results even in the presence of object motion. The presented approach (which we call the "Butterfly" method) involves reconstructing two initial volumes, each of which applies a higher view weight to the data at a different end of the short-scan. Although the two reconstructions have severe streaking artifacts, the direction of these artifacts is different between the two. A simple Fourier transform based blending of these two initial image volumes produces a final image volume with substantially reduced artifacts. Results are shown for simulated axial data (with a cone angle of 14.6 degrees) of the XCAT phantom.**

I. Introduction

Short-scan image reconstruction is an important problem since it is often not possible—or not desirable—to acquire a full 360 degrees of data in axial tomography. For example, C-arm systems often cannot complete a full-scan rotation due to mechanical limitations. Also, in some cases, it is useful to limit the data acquisition time as much as possible to reduce motion artifacts (e.g., for cardiac CT). Unfortunately, traditional short-scan reconstruction algorithms can produce images with an unacceptable level of cone-beam artifacts when the cone-angle is large enough to cover an entire organ in a single scan. The purpose of this paper is to describe an algorithm that produces images from such an acquisition that are dramatically better than those produced by a traditional short-scan FBP-type algorithm. The algorithm achieves this goal without significantly affecting noise or resolution and without altering the influence of object motion on the images.

We begin by reviewing three challenges associated with wide cone-angle axial reconstruction. First, even when acquiring a full 360 degree scan, it is well-known that there is missing frequency data due to fact that the axial trajectory does not satisfy Tuy's condition [1] for a 3D volume. Specifically, there are planes that intersect our 3D volume that do not intersect the source trajectory. This is true even when the detector is large enough to cover the entire object from every view. The result is that the reconstructed object ends up being similar to what one would get by applying a shift variant filter to the original image. The filter has a cone of missing frequencies in Fourier space, where the size/shape/orientation of this cone changes depending on where you look in the imaging volume [2]. Within the clinically relevant range of cone/fan angles, the percentage of the frequency data that is missing grows roughly in a quadratic manner with distance from the scan plane, and does not reach 1 percent of the total Fourier data for any points in the volume until the cone angle exceeds 14 degrees. The method described herein does not address this first problem.

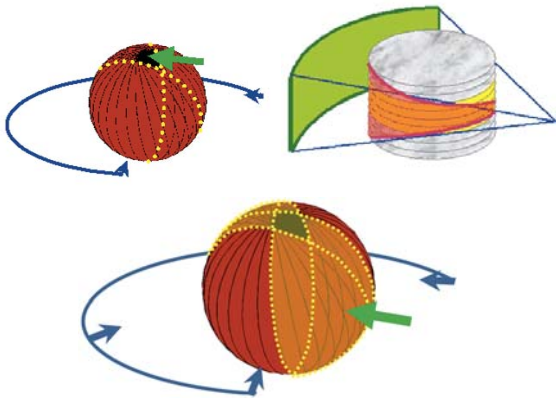


Fig. 1. Illustration of the three challenges in axial cone-beam reconstruction: i) missing data, ii) z-truncation, iii) mishandled data.

A second problem (that of z-truncation) arises when one attempts to reconstruct voxels that project outside the detector for some views, or when one uses a reconstruction method that requires filtering that extends to rays that are unmeasured due to the limits of the detector size. In this paper, we will focus primarily on the portion of the volume that always projects on the detector, and we will use 1D filtering in the projection domain that is oriented only along the rows of the detector. As such, we do not need see a need to face this problem directly in this work.

A third problem in axial cone-beam tomography is the handling of redundantly measured data. Traditionally, this is done through a view-weighting that is adopted from 2D algorithms, which can lead to significant mishandling of the data. For example, in an effort to reduce the time window associated with the data that contributes to the reconstruction

of each voxel, a binary (or nearly binary) view weighting function can be used. This is done by selecting for each voxel a segment of the axial scan that is centered on the motion phase of interest and is just large enough to produce rays through the voxel whose orientations span a full 180 degrees in xy. For example, segments AC or BD each provide 180 degrees of data for a point in the center of figure 3. In this case, there is a wedge of frequencies (containing as a subset the missing cone spoken of above) that is either unaccounted for or (equally problematic) over-weighted by a factor of two in the reconstruction . The percentage of the frequency data affected by this problem is larger than that affected by problem 1, and grows roughly linearly with distance from the scan plane (for clinically relevant distances). The percentage reaches 1 percent for a coverage of about two degrees, and reaches nearly 9 percent at the edge of a 14 degree cone angle system. Often, reconstructions make use of more than 180 degrees of parallel data and use a smoother (less binary) view weighting function to handle the redundancy. The view-weighting function is subject to a normalization condition that forces rays that are conjugate to one another in xy to have weights that sum to one. As indicated in [4], such an approach produces an image that is approximately a linear combination of a series of reconstructions (each using a slightly shifted 180 degrees subset of parallel data for any given voxel). This angular smoothing does not eliminate the artifacts associated with the mishandled data, but does disperse (blur) them over a larger area such that they are somewhat less problematic.

We now turn our attention to how this third problem (data mishandling) has been addressed in the past. One approach is to patch in the low frequencies from a full-scan by taking a difference between the short- and full-scan images and smoothing it before adding it back to the short-scan image [3]. This helps some, but leaves high-frequency artifacts uncorrected and reduces the temporal resolution of low-frequencies. An approach to handle all available data correctly is to use three families of tilted filtering lines as described in [5]. This approach is quite effective and even works fairly well in the presence of data truncation in spite of the fact that the filtering lines often extend off the edges of the detector in the z-direction. Apart from the computational challenges, however, this approach may also introduce streaky motion artifacts due to the discontinuous behavior of the effective frequency weighting. Another approach for dealing with this redundancy that has a similar set of advantages and disadvantages is given in [6]. Yet another approach [4] was designed to approximately cancel out the mishandling errors described above without affecting the noise or temporal behavior, but its effectiveness degrades for frequencies that are close to the cone of missing frequencies. Finally, there is an approach described in [7] that introduces the concept of combining multiple suboptimal reconstructions in a strategic way so as to leverage the good properties of each.

II. Method

We consider now a new approach (which we call the "Butterfly" method) that draws from several ideas introduced in the approaches mentioned above. The main new idea is to deal with data redundancy in the 2D Fourier space of the reconstructed images rather than by attempting to solve the redundancy through either view weighting, filtering in the projection domain, or object dependent image blending. We first produce two image volumes that are complementary to each other in the 2D Fourier space and then blend them together with filters. The Butterfly algorithm operates as follows for a cylindrical detector:

1) Data are rebinned to cone-parallel geometry and each row is weighted by the cosine of the cone angle.
2) A first view-weighting function is applied to the data.
3) The data set is filtered with a ramp filter.
4) Backprojection is performed to produce a first volume.
5) Steps 2-4 are repeated with a second view weighting function to produce a second volume.
6) For each slice of the volume, the final image (F) is produced from the two initial images (I_1 and I_2) by blending in 2D Fourier space as follows: $F = \text{IFFT}_{2d}\left[W\text{FFT}_{2d}(I_1) + (1-W)\text{FFT}_{2d}(I_2)\right]$, where W is a (smooth) weighting function.

Note that rebinning each detector row is done independently of other rows. Also note that the view-weighting step and the ramp filtering step are often seen in the opposite order as what we show here. Doing the view-weighting first limits the number of views that are needed for reconstruction. At the same time, the validity of the algorithm is unaffected (in both cases, the algorithm is "exact" in the central plane) since we are in parallel beam geometry.[1]

The key motivation for this approach can be appreciated by analyzing the frequency content of each of the two reconstructions. We start by noting that the ramp filter is applied in a direction that is tangent to the x-ray source path. This means that we can simply count the number of intersections between a Radon plane and the portion of the source path that is used to reconstruct a particular voxel in order to determine the effective weighting of the associated Fourier data. This is in contrast to methods like [5] and [6] which use Hilbert filters that operate in a different direction from the derivative in order to give some frequency contributions a negative weight. We can then make a map of this weight on a sphere in Frequency space since the weight depends only on direction.

We can build the frequency weighting map by considering a disk rotating in space. Each view contributes frequency

[1]The justification for this can be found by analyzing the influence of a delta function in the image domain at one point (P_1) in the scan plane on the reconstructed density at another point in the scan plane (P_2). This contribution can be further broken down into the contribution from each angle, where conjugate views are considered together. It can be shown that the sum of the two conjugate contributions are identical even if the individual contributions are not always identical. In one case, the weighting is based on the view weighting that P_1 sees, whereas in the other the weighting is based on the view weighting that P_2 sees and both view weighting functions meet the same normalization condition.

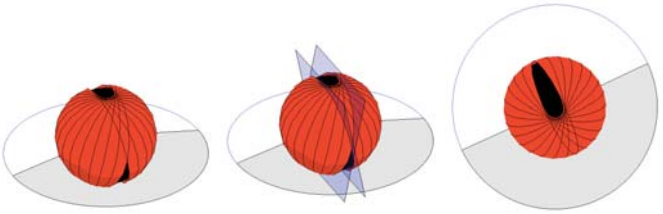


Fig. 2. Illustration of how Fourier space is swept out by a disk orthogonal to the ray direction for a source arc of length 180 degrees. Proceeding left-to-right, we have a perspective view of the coverage; the "corrupted" data wedge; and a view from the $+z$ frequency direction.

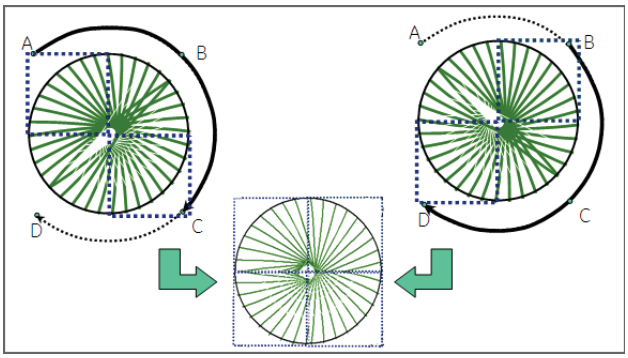


Fig. 3. The two initial reconstructions are reconstructed from source arcs AC and BD (each of which provides 180 degrees of data from the perspective of the voxel). Combining the best sections of frequency data in these two reconstructions nearly eliminates the mishandled frequencies.

information on a disk that is orthogonal to the ray direction. As described in [4], for a voxel of interest in the scan plane, this disk contains the axis about which it rotates (like a coin spinning very fast on a table). For a voxel of interest away from the scan plane, the disk is tilted such that it no longer contains the z-frequency axis. As it spins, it wobbles like a coin that is spinning *slowly* on a table. When all views that cover a 180 degree range from the perspective of the voxel are considered, this disk rotates and covers most of a sphere as shown in figure 2. There are some points within the sphere that are hit by the disk twice (notice the region where there is overlap between the disks), and some that are never hit by the disk (in the figure you can see a black sphere inside the red shell at these locations). Together, these points form a (double) wedge in frequency space of "corrupted" data (this wedge is shown in blue in the central panel of figure 2). We call this data corrupted, since it is accounted for either 0 or 2 times, while the remainder of the frequencies are weighted uniformly by 1. If all frequencies were weighted uniformly, we could have an "exact" reconstruction.

The wedge of "corrupted" Fourier data produces streaky cone beam artifacts that are oriented more or less along the direction connecting the endpoints of the segment of the source trajectory used in the reconstruction. As long as this direction is significantly different for the two reconstructions, the two wedges will have a small overlap in the frequency domain. We can use the amount of overlap between any given vertical plane in 3D Fourier space (which corresponds to a line in 2D Fourier space) and the wedge of mishandled frequencies

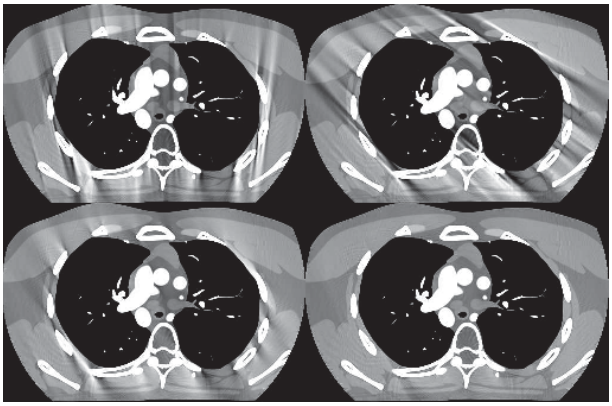


Fig. 4. XCAT phantom reconstructions (WW = 400 HU). top: two initial images of slice at 75 percent of maximum cone angle, bottom left: traditional Parker-FDK reconstruction , bottom right: Fourier combined (final) Butterfly reconstruction .

Fig. 5. Example weighting functions (W) that can be used in the 2D Fourier domain (white is 1, black is 0).

for the two different reconstructions as a way to select which reconstruction has less corruption for that particular frequency plane. This allows us to select the good regions from each reconstruction by means of a simple 2D filter.

An illustration of the effect of this process in the frequency domain is given in figure 3. One can appreciate that the corrupted wedge is mostly contained within quadrants 1 and 3 for the first initial reconstruction and is mostly contained within quadrants 2 and 4 in the second initial reconstruction. The result of the Fourier blending operation shows very little corrupted data.

It turns out that in practice, it is best to use a weighting function (W) that changes more gradually in the angular direction than the one used for illustrative purposes above. This helps to mitigate streaks caused by object motion. In figure 5, we show two sample weighting functions, one which changes abruptly in the angular direction and one that exhibits angular smoothing in order to mitigate artifacts from inconsistencies between conjugate ray measurements arising due to motion, scatter, etc.

We now discuss the view-weighting functions used for the two initial reconstructions. Since these view-weighting functions are 2D (independent of row), they can be created by considering only rays in the central plane. Note that for a partial scan, there are three types of such rays: i) those that have conjugates within the scan at a later view, ii) those that do not have conjugates, and iii) those that have conjugates

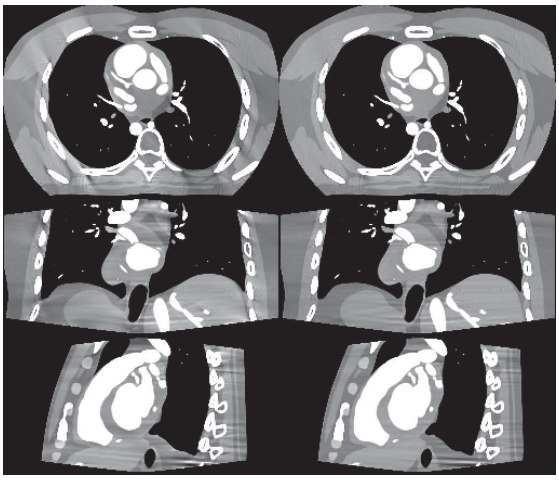


Fig. 6. XCAT reconstructions (WW = 400 HU) top: slice at 44 percent of max cone angle; mid: y = 18mm; bottom: x = 18mm; left: Traditional short-scan FDK; right: Butterfly method.

within the scan at an earlier view. The three arcs (AB, BC, and CD) in figure 3 have rays that correspond respectively to these three types for a voxel located on the rotation axis (assuming a clockwise source rotation). The goal in selecting these view weighting functions is to make the orientation of artifacts in the two image volumes as different as possible, so as to minimize the amount of overlap between the two "corrupted" data regions (see, e.g., the artifacts in the top two panels of figure 4). If exactly a short-scan of views is available, one of the two view weighting functions will have values of one for rays of type (i) and (ii) and will have values of zero for type (iii) rays. The other function will be one for rays of types (ii) and (iii) and zero for type (i) rays. This will ensure that the first reconstruction volume always uses the data that is as early as possible in the scan, while the other reconstruction always uses data that is as late in the scan as possible. If, however, the acquired data is much larger than the minimum scan range required for recon (e.g., we have nearly a full scan), this approach fails to provide maximum separation between the orientation of the artifacts. One solution is to limit the angular range for each voxel to 270 degrees of parallel data. In the case where the acquired data range is at least 270 degrees plus the full fan angle, this produces a very simple view weighting function in the parallel sinogram domain that depends only on angle: of the 270 degrees of parallel data, the first 180 degrees produces one reconstruction, while the last 180 degrees produces the other reconstruction. To minimize backprojection effort, the three 90 degree chunks can be backprojected separately and then combined with the central volume being taken as is and the other two filtered by W and $1 - W$.

III. Simulations/Results

In figures 4 and 6, we show results that illustrate the artifact reduction provided by the Butterfly method. Data for the XCAT phantom [8] was simulated with CatSim [9] in the cardiac region. The number of views and columns simulated was 984 and 888 respectively. The detector had a cone angle of 14.6 degrees. Oversampling of the detector, source, and view were performed. Reconstructions were performed with both a traditional FDK-type algorithm with Parker-like view weighting and the method proposed here.

IV. Discussion/Conclusions

It should be noted that the orientation of the "corrupted" wedge of frequency data will be somewhat shift-variant in the case where the view range of data is insufficient to produce 270 degrees of rebinned data. One could conceivably use shift-variant image filtering in response, but we have found that this complexity is usually not necessary in practice since the amount of corrupted data in the final recon is not very sensitive to small changes in the image filter direction.

We wish to point out that the Butterfly approach can also be used in the native geometry. In the cone-parallel geometry, the ramp filter is applied across the rebinned rows, so the direction of the filter is effectively always parallel to the vertex path tangent. Note that a slight approximation is made here which arises from the fact that our filtering effectively takes place on a curved surface rather than a plane, but this is not significant. In the native geometry, this approximation is not necessary. In this case, steps 1–4 are replaced by: 1) filtering, 2) view-weighting (and cosine-cone-weighting), 3) backprojection with a distance dependent weight.

The Butterfly method can produce high quality images from wide cone partial scan data. This capability is critical to the clinical success of a commercial wide-cone CT scanner. The Butterfly method performs very well at correctly handling the acquired frequencies. Although there are still some artifacts due to missing data (as seen in the reformatted slices), the images are far superior to those produced by a traditional short-scan algorithm. Remaining artifacts can be addressed by other means.

References

[1] Tuy H K, "An inverse formula for cone-beam reconstruction", *SIAM J. Appl. Math*, 43, 546-552, 1983.

[2] Bartolac S, Clackdoyle R, Noo F, Siewerdsen J, Moseley D, and Jaffray D, "A local shift-variant fourier model and experimental validation of circular cone-beam computed tomography artifacts", *Med. Phys.*, 2009.

[3] Horiuchi T, Nakanishi S, Akino N, "X-ray Computed Tomography Apparatus," U.S. Patent 7,751,524, Jul 2010.

[4] Pack J D, Manak J, "A novel approach to reconstruct cone-beam data acquired on less than a full scan", *Nuclear Science Symposium Conference Record*, 4229-4231, 2008.

[5] Nett B E, Zhuang T-L, Leng S, Chen G-H, "Arc based cone-beam reconstruction algorithm using an equal weighting scheme", *Journal of X-Ray Science and Technology*, 15, 19-48, 2007.

[6] Zhu L, and Yoon S, Fahrig R, "A short-scan reconstruction for cone-beam CT using shift-invariant FBP and equal weighting", *Med. Phys.*, 34 (11), 4422-4438, 2007.

[7] Dennerlein F, Noo F, Harer W, Hornegger J, Laurisch G, "Constriction of cone-beam artifacts by the z-smart reconstruction method", *Nuclear Science Symposium Conference Record*, 4090-4096, 2007.

[8] Segars W P, Tsui B M W, Frey E C, Fishman E K, "Extension of the 4D NCAT phantom to dynamic X-ray CT simulation," *Nuclear Science Symposium Conference Record*, 5, 3195, 2003.

[9] De Man B, Basu S, Edic P, Iatrou M, Tower B, Chandra N, Dunham B, McOlash S, Sainath P, Shaughnessy C, Williams E, "CatSim: a new computer assisted tomography simulation environment", *SPIE Medical Imaging*, 2007.

Fast Computation of Projections from Triangulated Surfaces

Stefan Sawall, Matthias Baer, Marcus Brehm, Michael Knaup, and Marc Kachelrieß

***Abstract*—The forward projection is an operation widely used in computed tomgraphy, e.g. in iterative reconstruction methods or the correction of prominent artifacts. In most cases voxels are used as basis functions within these algorithms. The construction of complex antropomorphic phantoms using voxels, however, is very time consuming. In most cases simpler shapes like ellipsoids are combined to construct complex objects. An alternative way to represent objects has been used in computer graphics for the past two decades: triangles. Especially computer aided design (CAD) software provides objects with nearly arbitrary complexity that might be used as phantoms or for artifact correction methods. However, the forward projection of such a triangulated object requires the calculation of all intersection points of a desired ray with all triangles of the object. As several billion rays have to be computed to obtain a complete CT dataset this method is only feasible for small objects. The authors therefore propose a fast method for the calculation of intersection lengths of highly complex objects containing millions of triangles. This method employs a spatial subdivision scheme to speed up the necessary computations. The results indicate that CT projection images of highly complex objects can be generated within seconds. This for example allows for the incorporation of known, triangulated objects in novel CT reconstruction and artifact correction methods.**

***Index Terms*—Computed Tomography, Forward Projection, Artifact Correction**

I. Introduction

THE forward projection is a common operation in computed tomography (CT). Applications include reconstruction algorithms [1, 2, 3], artifact correction methods [4] or calibration procedures [5]. Hence, several algorithms have been proposed in the literature to address this topic, e.g. Joseph's method [6] or Blobs [7]. All of these algorithms compute projection images based on objects represented by certain basis functions like voxels and Kaiser–Bessel functions, respectively. Altough computationally highly efficient, the creation of complex, antropomorphic phantoms using these object representations is a difficult and time consuming task as these phantoms are often constructed using only basic shapes, e.g. ellipsoids and cylinders, and constructive solid geometry (CSG) operations [8]. An alternate way of defining such phantoms is provided by computer aided design (CAD) software that allows to model arbitrary objects represented by spline surfaces. These surfaces can be sampled with almost arbitrary precision resulting in a point cloud which can be triangulated using e.g. the marching cubes algorithm [9] or Delauney triangulation [10]. Such an object represented by triangles can be used to calculate projection images required for image reconstruction or artifact correction. Prominent applications of this approach other than phantom simulations include the reconstruction of objects with known components [11] or scatter correction methods in dimensional CT [12] where CAD models of the objects are provided.

The authors herein propose a method for the fast forward projection of triangulated surfaces to minimize the time required for the computation of CT projection images of highly complex objects. This problem corresponds to the computation of the intersection points of a ray originating at the x–ray source heading towards a detector pixel with all triangles of the object. To speed up this process we propose to use a spatial subdivision structure to reduce the number of triangles that have to be considered per ray. This allows for the computation of projection images of highly complex objects containing millions of triangles within only a few seconds.

Dr. Stefan Sawall, Dr. Matthias Baer, Marcus Brehm, Dr. Michael Knaup, and Prof. Dr. Marc Kachelrieß: Medical Physics in Radiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

Dr. Stefan Sawall, Marcus Brehm, and Prof. Dr. Marc Kachelrieß: Institute of Medical Physics, Friedrich–Alexander–University (FAU) Erlangen–Nürnberg, 91052 Erlangen, Germany.

Corresponding author: stefan.sawall@dkfz.de.

II. Materials and Methods

A. Overview

The computation of CT projection images from triangulated surfaces can be divided into the following three steps:

- Generation of the triangulated surfaces
- Computation of intersection lengths for all rays with these surfaces
- Weighting of intersection lengths with attenuation coefficients to obtain projection images

We herein assume that the triangulated surfaces are provided and will not further consider their generation. Additional details can for example be found in reference [13]. We further assume that the provided objects are watertight, i.e. that the triangles associated to these objects provide closed surfaces without any gaps.

The calculation of intersection lengths of a ray originating at a source position heading towards a certain position on the detector with an object requires the computation of the intersection points between the ray and the triangles of this object. Note that the source is not necessarily an x–ray tube. Rays are also allowed to start within objects enabling the simulation of positron emission tomography (PET) or single photon emission computed tomography (SPECT) systems. The fact that a densely sampled object may contain several million triangles prohibits the calculation of the intersection points of all triangles with every simulated ray. A typical CT dataset requires the simulation of several billion rays and is therefore

computationally highly demanding. To overcome this issue we propose to use a spatial subdivision structure to speed up these computations. In particular, we use an octree described in section II-B to reduce the number of ray–triangle intersection tests that have to be performed per ray. The result of the remaining intersection tests is a sorted list of intersection points with respect to the source position for a given ray with an object. The used computations (see section II-C for details) further allow to determine if the ray enters or leaves the object. Thus, the intersection lengths can be computed by summing the lengths between the intersection points corresponding to an entrance into the object and the adjacent points corresponding to an exit.

A subsequent step allows for a weighting of the obtained intersection lengths with attenuation coefficients of the desired materials to obtain monochromatic projection images or the usage of the intersection lengths in a polychromatic model to include, e.g., beam hardening effects [14].

B. Spatial Subdivision

As it is not computationally reasonable to calculate the intersection points of all triangles with all simulated rays, a spatial subdivision structure is used to speed up these computations. In particular an octree is used in the following. An octree in three-dimensional space is a tree structure with each node in general containing eight child nodes. If we consider the root node to be the bounding box of our triangulated objects these child nodes correspond to equally sized spatial subdivisions, i.e. rectangular boxes, of this bounding box. These nodes will be referred to as internal nodes in the following. This subdivision process continues by subdividing each child node in eight boxes again. If the space enclosed by such a subbox does not contain any triangles, it will no further be used to spawn new child nodes. I.e. its parent node rejects this child node and thus the number of childs is reduced by one. As soon as certain termination criteria are met the process stops. The subdivision in our case stops as soon as no new subboxes can be generated that contain at least 16 triangles. The triangles contained in a subbox are stored in a so–called external child node. Note that several highly performant algorithms for the creation of this data structure exist [15].

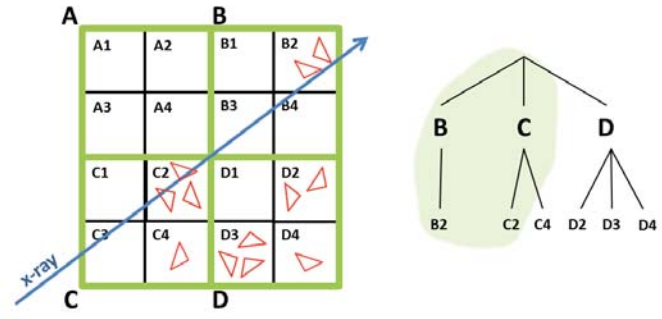


Fig. 1. Left: The bounding boxes in subdivision level 1 (green) and subdivision level 2 (black) to sort the red triangles into an octree. Right: The created octree and the considered nodes (green) for the ray shown left.

If the intersections of a ray with an object shall be computed the octree is traversed. I.e., it is ensured that the ray intersects with the bounding box of the root node. If this is not the case the procedure is terminated. If the ray intersects the root node all internal child nodes are recursively checked for an intersection. If any of these nodes contains an external child node the intersection points with the triangles enclosed therein are computed. Using this spatial subdivision scheme ensures that the number of triangles that have to considered per ray is highly reduced and thus performance is increased.

This approach is illustrated in figure 1 for a two–dimensional problem. The left part of figure 1 schematically shows the created subboxes corresponding to some triangles marked in red. As we only consider a two–dimensional problem for illustration purposes the first subdivision of the bounding box creates four subboxes denoted with ABCD. The next level of subdivisions partitions these boxes in A_i, B_i, C_i and D_i with $1 \leq i \leq 4$. However, the branch of boxes A_i is not included in the octree shown on the right hand side of the figure, as box A does not contain any triangles. A ray traversing this tree indicated in blue on the left thus only has to consider the objects located in the boxes C_2 and B_2. This is illustrated in green on the right hand side of the figure. All other triangles are neglected as their corresponding nodes or boxes, respectively, are never hit.

C. Ray–Triangle Intersection

The intersection of a ray with a triangle can be computed using several methods, e.g. including Plücker and barycentric coordinates [16]. We use a variant of the method proposed in reference [17]. This method briefly works as follows. We assume a triangle defined by its three vertices $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^3$. First of all a test is performed that ensures that a desired ray $\boldsymbol{r} = \boldsymbol{s} + \lambda \boldsymbol{t}$ with $\boldsymbol{s} \in \mathbb{R}^3$ being the source position and heading in direction of $\boldsymbol{t} \in \mathbb{R}^3$ intersects with the plane spanned by the vertices of the triangle. If this tests fails the current intersection routine is aborted and the algorithm proceeds with the next triangle. If the ray hits the plane several tests are performed to ensure that the intersection point is within the boundaries of the triangle. As a result the parameter $\lambda \in \mathbb{R}$ is obtained as well as an information whether the ray leaves or enters the object. The latter can easily be determined using the normal vector of the triangle. These information are sorted with respect to λ and are used in a subsequent step to compute intersection lengths by summing the distances between consecutive entry– and exit–points.

D. Test Configuration

To illustrate the capabilities of the proposed method two test objects have been designed. Both objects are procedurally generated and allow to dynamically increase the number of triangles. Object one is a sphere with a diameter of 35 mm located in the center of rotation. Each point on the surface of the sphere was perturbed in radial direction using a uniform distribution with a maximum deviation of 1 mm to simulate surface roughness. Object two consists of a $3 \times 3 \times 3$ grid of equidistantly spaced spheres, each with a diameter of 15 mm,

TABLE I: Total runtime of the algorithm in seconds using the test cases described in section II-D. Runtimes only include the forward projection itself. Time to load the data and to create the octree is not included.

	Test object 1						Test object 2					
Cores \ Triangles	2500	$1 \cdot 10^4$	$5 \cdot 10^4$	$5 \cdot 10^5$	$1 \cdot 10^6$	$25 \cdot 10^6$	2500	$1 \cdot 10^4$	$5 \cdot 10^4$	$5 \cdot 10^5$	$1 \cdot 10^6$	$25 \cdot 10^6$
1	246 s	309 s	355 s	529 s	610 s	3474 s	500 s	790 s	1040 s	1481 s	1610 s	3088 s
4	63 s	80 s	91 s	136 s	158 s	910 s	125 s	205 s	265 s	375 s	412 s	776 s
8	32 s	41 s	47 s	70 s	82 s	468 s	65 s	103 s	135 s	194 s	212 s	426 s
12	22 s	28 s	32 s	47 s	55 s	315 s	44 s	70 s	91 s	131 s	143 s	285 s

TABLE II: TCMR values in seconds for all test cases.

	Test object 1						Test object 2					
Cores \ Triangles	2500	10^4	$5 \cdot 10^4$	$5 \cdot 10^5$	$1 \cdot 10^6$	$25 \cdot 10^6$	2500	10^4	$5 \cdot 10^4$	$5 \cdot 10^5$	$1 \cdot 10^6$	$25 \cdot 10^6$
1	0.34 s	0.43 s	0.49 s	0.73 s	0.85 s	4.83 s	0.69 s	1.10 s	1.44 s	2.06 s	2.24 s	4.29 s
4	0.35 s	0.44 s	0.51 s	0.76 s	0.88 s	5.06 s	0.69 s	1.14 s	1.47 s	2.08 s	2.29 s	4.31 s
8	0.36 s	0.46 s	0.52 s	0.78 s	0.91 s	5.20 s	0.72 s	1.14 s	1.50 s	2.16 s	2.36 s	4.73 s
12	0.37 s	0.47 s	0.53 s	0.78 s	0.92 s	5.25 s	0.73 s	1.17 s	1.52 s	2.18 s	2.38 s	4.75 s

to illustrate the performance in the presence of multiple objects. The number of triangles per test object was varied between 2500 and $25 \cdot 10^6$.

The simulated geometry provides a source–isocenter–distance of 572 mm and a isocenter–detector–distance of 375 mm. A flat detector with a matrix dimension of 1000×1000 pixels, each of size 388 µm, is used. A dataset consists of 720 projections equidistantly spaced in angular direction over $360°$. Thus, a complete dataset requires the intersection of 720×1000^2 rays with the test objects. I.e. in case of $25 \cdot 10^6$ triangles per test object this results in about $1.8 \cdot 10^{15}$ ray–triangle intersection tests for a simulated dataset.

To objectively quantify the performance of the method the time required per 2^{20} simulated rays and CPU cores (TCMR) for a given number of triangles per object is evaluated as

$$TCMR = \frac{t \cdot N_C}{MR}. \quad (1)$$

Therein MR is the number of simulated rays divided by 2^{20} to obtain megarays. This is a constant of 720 in all tests due to the number of projections and their size used. The number of used CPU cores is denoted as N_C and the time required to complete a certain test, i.e. to create 720 projection images each of size 1000×1000, is denoted with t measured in seconds. This measure allows to quantify the dependency of the algorithm on the number of triangles per object and possible memory bootlenecks in the case of a parallel processing. I.e., if the algorithm is not limited by memory bandwidth TCMR is independent of the number of used CPU cores for a given number of triangles.

All tests are conducted using single precision floating–point arithmetic in a CPU implementation of the proposed method. The system running the tests provides two Intel® Xeon X5690 hexacore processor, i.e. 12 physical processor cores with 3.46 GHz each, and 96 GB of RAM. Hyperthreading was disabled during all tests.

III. Results

Figure 2 shows intersection lengths created by the proposed method for the two test cases. These data were used to reconstruct images assuming monochromatic radiation and objects made of water. An evaluation of the CT values inside the spheres resulted in a mean CT–value of 0 HU with a standard deviation of about 0.3 HU.

Table I shows the runtime in seconds of the proposed method for both test cases, different triangle counts per test case and different numbers of CPU cores used to parallelize the computations. Note that the creation of the octree is not included in these measurements. In the worst case of $25 \cdot 10^6$ triangles per test case a complete dataset can be obtained in less than 3500 seconds in all cases. Note that typical datasets only contain one to two million triangles, resulting in a complete dataset within less than a minute. Altough both test cases provide the same triangle counts the total runtime differs significantly. E.g. using a single processor and $5 \cdot 10^5$ triangles results in a temporal demand of 529 s for test case 1 and 1481 s for test case 2. This is caused by the fact that the octree contains many more nodes if multiple, spatially disjunct objects are used and thus the time for tree traversal is increased. In general the proposed, CPU–based method is faster than other methods recently proposed in the literature. Reference [18] for example reports the temporal demand for the forward projection of $4.6 \cdot 10^5$ triangles onto a detector of size 640×480 using a GPU–based method as 173 ms per projection. If we assume that this method scales linear with the number of simulated rays and the number of triangles, the time to compute a projection of size 1000^2 from $5 \cdot 10^5$ triangles would roughly be 612 ms. This is more than three times slower than the corresponding value of 188 ms per projection (136 s for a complete dataset) reported in table I using only four CPU–cores.

The TCMR values obtained from the benchmarks are shown in table II. The fact that these values for a given test object and a given number of triangles do not significantly vary with the number of used CPU cores indicates that the proposed methods is not bandwidth–limited and scales linear with the number of processors used.

IV. Conclusion and Discussion

The authors proposed a method for the computation of intersection lengths from triangulated surfaces. The algorithm

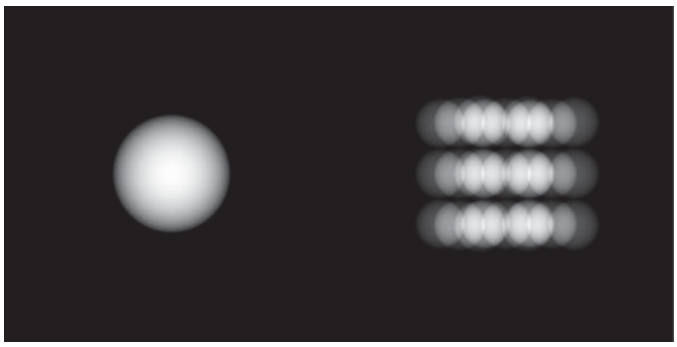


Fig. 2. Intersection lengths computed for test case 1 (left) and test case 2 (right).

employs a spatial subdivision scheme to speed up the necessary calculations. It was possible to demonstrate that intersection lengths of objects containing several million triangles can be calculated in a few seconds using modern CPUs. The proposed algorithm scales linear with the number of used processor cores and thus provides highest performance on server platforms, e.g. for the computation of projections from semi–antropomorphic phantoms [19]. This enables the usage of highly complex, triangulated objects for the simulation, image reconstruction and artifact correction in computed tomography and other imaging modalities.

Acknowledgment

We thank the Intel Corporation for providing their latest multi–core platforms. The reconstruction software RayConStruct–IR was provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrasonic Imaging*, vol. 6, pp. 81–94, 1984.

[2] F. J. Beekman and C. Kamphuis, "Ordered subset reconstruction for x–ray CT," *Physics in Medicine and Biology*, vol. 46, no. 7, pp. 1835–1844, 2001.

[3] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Physics in Medicine and Biology*, vol. 56, no. 6, p. 1545, 2011.

[4] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.

[5] A. M. Kingston, A. Sakellariou, A. P. Sheppard, T. K. Varslot, and S. J. Latham, "An auto-focus method for generating sharp 3D tomographic images," in *Proc. of the SPIE, Developments in X-Ray Tomography VII*, vol. 7804, no. 1. SPIE, 2010, pp. 78 040J–78 040J–15.

[6] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Transactions on Medical Imaging*, vol. MI–2, no. 3, pp. 192–196, 1982.

[7] H. Wang, L. Desbat, and S. Legoupil, "Image representation by blob and CT reconstruction from few projections," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) Record*, 2011, pp. 3971 –3976.

[8] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch, and J. Hornegger, "Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom," *Physics in Medicine and Biology*, vol. 57, no. 13, pp. N237–N252, 2012.

[9] W. E. Lorensen and H. E. Cline, "Marching Cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.

[10] L. Guibas, D. Knuth, and M. Sharir, "Randomized incremental construction of delaunay and voronoi diagrams," *Algorithmica*, vol. 7, pp. 381–413, 1992.

[11] J. W. Stayman, Y. Otake, A. Uneri, J. L. Prince, and J. H. Siewerdsen, "Likelihood-based CT reconstruction of objects containing known components," in *Proceedings of the 11th International Meeting on Fully 3D Image Reconstruction and the 3rd Workshop on High Performance Image Reconstruction*, 2011, pp. 254–7.

[12] M. Baer, M. Hammer, M. Knaup, I. Schmidt, R. Christoph, and M. Kachelrieß, "Scatter correction methods in dimensional CT," in *Conference on Industrial Computed Tomography 2012*, 2012.

[13] W. Cho, T. Maekawa, N. Patrikalakis, and J. Peraire, "Robust tessellation of trimmed rational B-spline surface patches," in *Proceedings of Computer Graphics International*, 1998, pp. 543 –555.

[14] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (EBHC) for CT," *Medical Physics*, vol. 37, no. 10, pp. 5179–5187, 2010.

[15] K. Zhou, M. Gong, X. Huang, and B. Guo, "Data–parallel octrees for surface reconstruction," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 5, pp. 669–681, 2011.

[16] M. Shevtsov, A. Soupikov, A. Kapustin, and N. Novorod, "Ray-Triangle Intersection Algorithm for Modern CPU Architectures," in *Proceedings of GraphiCon'2007*, 2007.

[17] T. Möller and B. Trumbore, "Fast, minimum storage ray-triangle intersection," *J. Graph. Tools*, vol. 2, no. 1, pp. 21–28, 1997.

[18] A. Maier, H. G. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast simulation of x-ray projections of spline-based surfaces using an append buffer," *Physics in Medicine and Biology*, vol. 57, no. 19, pp. 6193–6210, 2012.

[19] C. Hoogendoorn, N. Duchateau, D. Sanchez-Quintana, T. Whitmarsh, F. Sukno, M. De Craene, K. Lekadir, and A. Frangi, "A high-resolution atlas and statistical model of the human heart from multislice ct," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 1, pp. 28 –44, 2013.

Angular Rebinning for Geometry Independent SPECT Reconstruction

Alexandre Bousse, Kjell Erlandsson, Stefano Pedemonte, Sébastien Ourselin, Simon Arridge, Brian F. Hutton

***Abstract*—This work proposes a novel approach to model the collimator response in SPECT. The approach consists of projecting the activity volume on a high number of virtual projection planes that are then averaged with an angular point spread function. It was motivated by the new possibilities offered by GPU for 3-D projection/backprojection. This approach also allows to model a wide range of SPECT imaging systems. Results show that reconstruction using our resolution modelling method is consistent with standard blurring. As an example, we show how to implement a convergent collimator response.**

***Index Terms*—SPECT reconstruction, resolution modelling, Hermitian adjoint**

I. Introduction

Single photon emission computed tomography (SPECT) imaging is a routine clinical procedure in nuclear medicine. Accurate image reconstruction requires a precise knowledge of the system matrix *i.e.* the probabilities that a photon emitted from a given position is detected at a given bin. This knowledge depends on several factors, such as the attenuation map [1], [2], the gamma camera geometry and septal penetration [3]. When the system matrix is known, the activity distribution image can be reconstructed by maximising the log-likelihood [4], [5] or penalised log-likelihood [6], [7]. In parallel hole SPECT, it is possible to efficiently project (resp. backproject) the activity distribution (resp. the sinogram) by convolving the activity volume slice by slice by a distant-dependent point spread function (PSF) [2]. This approach usually requires the assumption that the attenuation map within the cone of detection corresponds to the attenuation along the central line.

For other imaging system geometries (convergent SPECT, multi-pinhole, etc.), the above approach is not always feasible and projecting/backprojecting requires the computation of the system matrix. This can be achieved by measuring point source responses [8], [9] or Monte-Carlo simulation [10]–[12]. Obviously, in addition to being unable to incorporate the patient-dependent attenuation map, these approaches are normally too time consuming to be performed on-line.

In this work we propose a 2 step projector that can model a wide range of SPECT imaging systems. The idea was suggested in [13]. The first step consists of projecting the activity distribution on a large number of "virtual" azimuthal and polar angles. This step is performed efficiently using the GPU-accelerated Matlab toolbox NiftyRec [14]. The second step, presented in section II, is a data rebinning operation that takes the form of an angular convolution. Its adjoint operator can be computed so that exact backprojection can be performed. The method does not make use of the central line approximation with respect to the attenuation. In section III two examples of angular PSF are presented: parallel hole and fan-beam geometry. Discussion and conclusion are given in section IV.

Manuscript received February 4, 2013; revised April 16, 2013.

A. Bousse and K. Erlandsson are supported by a centre grant jointly funded by Cancer Research UK (CRUK) and EPSRC. UCLH/UCL receives a proportion of its funding from the UK Department of Health's NIHR Biomedical Research Centre's funding scheme.

A. Bousse, K. Erlandsson and B. Hutton are with Institute of Nuclear Medicine–UCL, University College London, London NW1 2BU, UK. B.F. Hutton is also with the Centre for Medical Radiation Physics at the University of Wollongong, NSW Australia.

S. Pedemonte, S. Ourselin and S. Arridge are with Centre for Medical Image Computing, University College London, London WC1E 6BT, UK.

email: a.bousse@ucl.ac.uk

II. Theory

Let $(\boldsymbol{o}, \overrightarrow{\boldsymbol{x}}, \overrightarrow{\boldsymbol{y}}, \overrightarrow{\boldsymbol{z}})$ be an orthonormal coordinate system in $\mathbb{R}^3$ and $\Omega \subset \mathbb{R}^3$ be the field of view. Without loss of generality we can assume Ω to be the unit ball. The activity distribution can be seen as a function $f(\boldsymbol{r})$ with $\boldsymbol{r} \in \Omega$. The operator $\mathcal{P}$ that maps f into the set of its line integrals is called the *X-ray transform* [15]. The choice of its parametrisation varies across the literature. For this work we define it as follows: let $\mathscr{P}(\varphi, \vartheta)$ be the plane tangent to the unit sphere $\partial\Omega$ at $\boldsymbol{o}(\varphi, \vartheta)$, the point of spherical coordinates $(1, \varphi, \vartheta)$ where $\varphi \in [0, 2\pi[$ is the azimuthal angle and $\vartheta \in [-\pi/2, \pi/2[$ is the polar angle. Let $(\boldsymbol{o}(\varphi, \vartheta), \overrightarrow{\boldsymbol{i}}(\varphi, \vartheta), \overrightarrow{\boldsymbol{j}}(\varphi, \vartheta))$ be a coordinate system on $\mathscr{P}(\varphi, \vartheta)$ such that it coincides with $(\boldsymbol{o}, \overrightarrow{\boldsymbol{x}}, \overrightarrow{\boldsymbol{z}})$ when $\varphi = \vartheta = 0$. The X-ray transform of f on $\mathscr{P}(\varphi, \vartheta)$ at position (x, y) is given by the line integral

$$\mathcal{P}f(x, y, \varphi, \vartheta) = \int_{-\infty}^{+\infty} f(\boldsymbol{p}_{\varphi,\vartheta}(x, y) + t\,\overrightarrow{\boldsymbol{d}}(\varphi, \vartheta))\,\mathrm{d}t \quad (1)$$

where $\overrightarrow{\boldsymbol{d}}(\varphi, \vartheta)$ is the unitary vector normal to $\mathscr{P}(\varphi, \vartheta)$ (pointing to the exterior of Ω) and $\boldsymbol{p}_{\varphi,\vartheta}(x, y)$ is the point of coordinates (x, y) on $\mathscr{P}(\varphi, \vartheta)$ in the $(\boldsymbol{o}(\varphi, \vartheta), \overrightarrow{\boldsymbol{i}}(\varphi, \vartheta), \overrightarrow{\boldsymbol{j}}(\varphi, \vartheta))$ coordinates system. When defined over a set of functions that are square-integrable, $\mathcal{P}$ is a continuous operator (see [15], p. 17). In SPECT imaging, the attenuation map $\mu(\boldsymbol{r})$ is incor-

porated by multiplying $f(\boldsymbol{p}_{\varphi,\vartheta}(x,y) + t\,\overrightarrow{\boldsymbol{d}}(\varphi,\vartheta))$ with $\exp\left(-\int_t^{+\infty}\mu(\boldsymbol{p}_{\varphi,\vartheta}(x,y)+t'\,\overrightarrow{\boldsymbol{d}}(\varphi,\vartheta))\,\mathrm{d}t'\right)$ in (1).

The idea developed here is to re-bin a complete line integral dataset $h(x,y,\varphi,\vartheta) = \mathcal{P}f(x,y,\varphi,\vartheta)$ to model a wide range of imaging systems. Continuous re-binning of $h(x,y,\varphi,\vartheta)$ takes the form of data re-blurring with some weighting function. Assume we wish to model a SPECT gamma camera that rotates around the $\overrightarrow{z}$-axis at a distance ρ to the origin and such that each projection plane is contained in a $[-1,1]^2$ square: the re-binned projection data are obtained by an operator $\mathcal{A}$ defined by

$$\begin{aligned}\mathcal{A}h(x,y,\varphi) &= \int_{-\pi/2}^{\pi/2}\int_0^{2\pi} h(u_{\varphi'}(x),u_{\vartheta'}(y),\varphi+\varphi',\vartheta') \\ &\times\ w(x,y,\varphi',\vartheta')\chi(x,y)\,\mathrm{d}\varphi'\mathrm{d}\vartheta' \qquad (2)\end{aligned}$$

where $u_\delta(t) = \rho\sin\delta + t\cos\delta$, $\chi(x,y)$ is the characteristic function of $[-1,1]^2$ and w is some weighting function. The angular blurring $\mathcal{A}$ can be easily interpreted: at camera position φ and detector bin location (x,y), $\mathcal{A}$ accounts for photons travelling in a direction defined by (φ',ϑ') with a contribution $w(x,y,\varphi,\vartheta)$. The two terms $u_{\varphi'}(x)$ and $u_{\vartheta'}(y)$ indicate where the corresponding photon trajectories are located in the complete dataset $h(x,y,\varphi,\vartheta)$, see figure 1. The geometry of the imaging system is determined by the choice of w. For example, if w does not depend on (x,y), $\mathcal{A}$ models a parallel hole collimator. The theoretical observation operator with angular blurring-based resolution modelling is $\mathcal{H} = \mathcal{A}\mathcal{P}$.

In order to utilise $\mathcal{A}$ for iterative image reconstruction, *i.e.* within an iterative algorithm [4]–[7], its Hermitian adjoint (transpose) must be computed.

Proposition 1. *Let $\mathcal{A}$ be as in (2) and denote $T_1 = [0,2\pi]$ and $T_2 = [0,2\pi]\times[-\pi/2,\pi/2]$. Let X (resp. Y) be the subset of $L^2(\mathbb{R}^2\times T_2)$ (resp. $L^2(\mathbb{R}^2\times T_1)$) composed of functions compactly supported on $[-1,1]^2\times T_2$ (resp. $\mathbb{R}^2\times T_1$). Assume there exists a function $K:\Omega\to\mathbb{R}^+$ such that for all $(x,y)\in[-1,1]^2$, $w(x,y,\varphi,\vartheta)\le K(\varphi,\vartheta)$ and*

$$\sup_{\vartheta,\varphi}\frac{K^2(\varphi,\vartheta)}{|\cos\vartheta\cos\varphi|} = C(w) < +\infty.$$

Then $\mathcal{A}$ is a bounded operator with $\|\mathcal{A}\|\le\pi^{3/2}\sqrt{2C(w)}$ and its adjoint operator $\mathcal{A}^*:Y\to X$ is given by

$$\begin{aligned}\forall g\in Y,\quad \mathcal{A}^*g(x,y,\varphi,\vartheta) &= \int_0^{2\pi} g(v_{\varphi-\varphi'}(x),v_\vartheta(y),\varphi') \\ &\times\frac{w(v_{\varphi-\varphi'}(x),v_\vartheta(y),\varphi-\varphi',\vartheta)}{|\cos(\varphi-\varphi')\cos\vartheta|}\chi(x,y)\,\mathrm{d}\varphi',\end{aligned}$$

where $v_\delta(t) = u_\delta^{-1}(t)$.

Proof. The adjoint is obtained by substituting (x,y) for $v_{\varphi'}(x)$ and $v_{\vartheta'}(y)$ when writing the adjoint equality $\int\mathcal{A}hg = \int h\mathcal{A}^*g$. The division by $|\cos\vartheta\cos\varphi|$ is a Jacobian. The same trick is used to find an upper bound for $\|\mathcal{A}\|$. (Full proof available on demand). □

The condition on w mean trajectories of the photons should not be parallel to the gamma-camera, which is always true. In practice the X-ray transform $\mathcal{P}$ is discretised to a $N\times m$ matrix $\boldsymbol{P} = \mathsf{D}_{N,m}(\mathcal{P})$ where $\mathsf{D}_{\cdot,\cdot}$ is a discretisation operator, and $\mathcal{A}$ is discretised to a $n\times N$ matrix $\boldsymbol{A} = \mathsf{D}_{n,N}(\mathcal{A})$. To perform the angular blurring (2) it is required that $N\gg n$. The full SPECT finite-dimensional projector is a $n\times m$ matrix $\boldsymbol{H} = \boldsymbol{A}\boldsymbol{P}$ and its transpose used for iterative reconstruction is $\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$. We utilised $\mathsf{D}_{m,N}(\mathcal{P}^*)\mathsf{D}_{N,n}(\mathcal{A}^*)$ to approximate $\boldsymbol{H}^{\mathrm{T}}$. In section III we demonstrate that $\boldsymbol{H}^{\mathrm{T}}\approx\mathsf{D}_{m,N}(\mathcal{P}^*)\mathsf{D}_{N,n}(\mathcal{A}^*)$. Resolution modelling utilising $\boldsymbol{H}$ and $\boldsymbol{H}^{\mathrm{T}}$ shall be referred to as *angular blurring projection* (ABP). Standard resolution modelling shall be referred to as *standard blurring projection* (SBP) implemented as in [2].

Because $N\gg n$ (*i.e.* high number of virtual projections), $\boldsymbol{P}$ and $\boldsymbol{P}^{\mathrm{T}}$ should be implemented efficiently. For this work we utilised the GPU-accelerated Matlab toolbox NiftyRec [14].

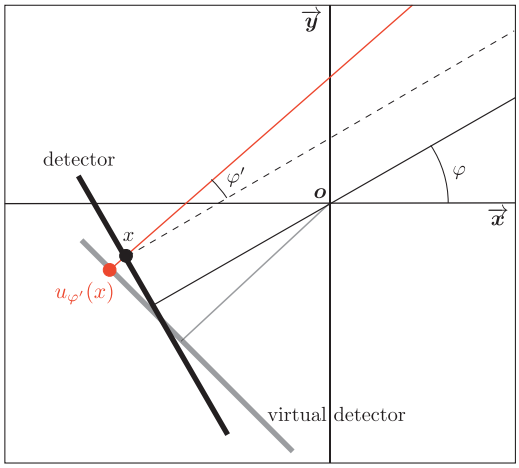


Fig. 1. Illustration of the angular re-binding: $u_{\varphi'}(x)$ is the virtual camera detection location of a photon hiting the true camera at angular position φ at location x with an angle φ'.

III. Results

A. Validation of the adjoint operator

In this section we experimentally verify that $\mathsf{D}_{m,M}(\mathcal{A}^*)\approx\boldsymbol{A}^{\mathrm{T}}$. This is can be considered as an experimental verification of proposition 1. For this purpose we randomly generate two sequences $\{\boldsymbol{u}_k\}$ and $\{\boldsymbol{v}_k\}$ where $\boldsymbol{u}_k\in\mathbb{R}^n$ and $\boldsymbol{v}_k\in\mathbb{R}^N$, $N\gg n$. For each k, we verify that

$$\langle\boldsymbol{A}\boldsymbol{u}_k,\boldsymbol{v}_k\rangle_{\mathbb{R}^N}\approx\langle\boldsymbol{u}_k,\mathsf{D}_{N,n}(\mathcal{A}^*)\boldsymbol{v}_k\rangle_{\mathbb{R}^n}. \qquad (3)$$

Figure 2 shows that (3) is a good approximation.

B. Imaging system examples

In this section we show 2 examples of SPECT systems: parallel hole and convergent cone-beam collimators. As briefly explained in section II, parallel hole collimators can be modelled with ABP using a position-independent PSF. Here we used a two-dimensional Gaussian PSF with diagonal covariance matrix *i.e.*

$$w_{\mathrm{par}}(\varphi,\vartheta)\propto\chi_R(\varphi,\vartheta)\exp(-\varphi^2/2\sigma_\varphi^2)\exp(-\vartheta^2/2\sigma_\vartheta^2) \qquad (4)$$

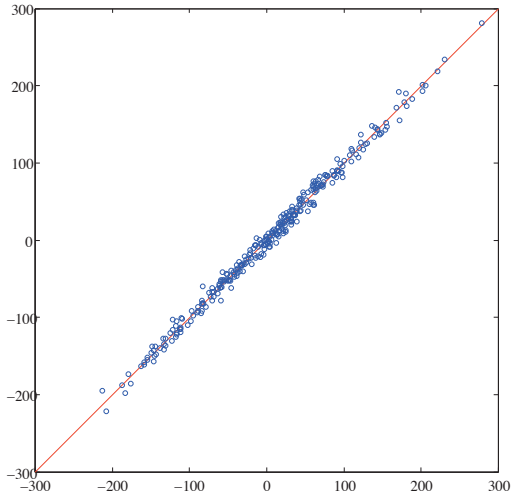
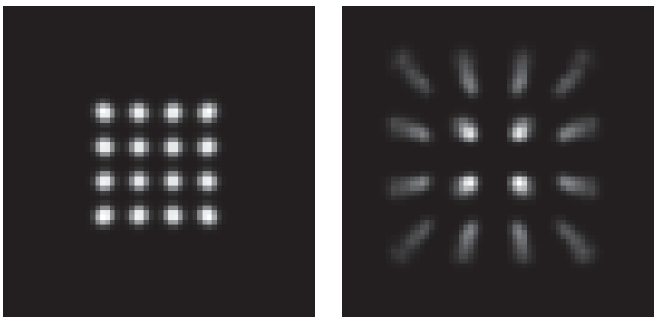


Fig. 2. Plot of 300 points whose coordinates are the left and right hand side of (3).



(a) w_{par} (b) w_{conv}

Fig. 3. Projection with resolution modelling following (2): (a) parallel hole collimator *i.e.* using (4); (b) convergent fan-beam collimator *i.e.* using (5)

where R is such that $d(R, \{-\pi/2, \pi/2\}) > 0$ and χ_R is the corresponding characteristic function. The presence of χ_R is necessary to ensure the hypothesis of proposition 1 is true *i.e.* by excluding angles $\pm\pi/2$. The convergent and divergent geometry PSF's are built upon w_{par} with the introduction of a term that changes the angular centring depending on the position (x, y) on the camera:

$$w_{\text{conv}}(x, y, \varphi, \vartheta) = w_{\text{par}}(\varphi + x\varphi_{\max}, \vartheta + y\vartheta_{\max}). \quad (5)$$

We chose a linearly-dependent position centring but other position dependencies can be used. Note that divergent geometries can be implemented by replacing $\varphi_{\max}$ and $\vartheta_{\max}$ with $-\varphi_{\max}$ and $-\vartheta_{\max}$ respectively. Figure 3 shows the results of noiseless projection using parallel hole (figure 3(a)) and convergent geometry (figure 3(b)). The projected phantom consists of 64 spheres distributed uniformly in a cube. The volume size is $m = 64^3$ and the projection data size is $n = 64^2 \times 120$. We used 360 virtual azimuthal angles φ and 360 polar angles ϑ to compute $\mathcal{A}$ (*i.e.* $N = 64^2 \times 360^2$). With parallel geometry only the first layer is visible, whereas using convergent geometry the 3 next layers are visible.

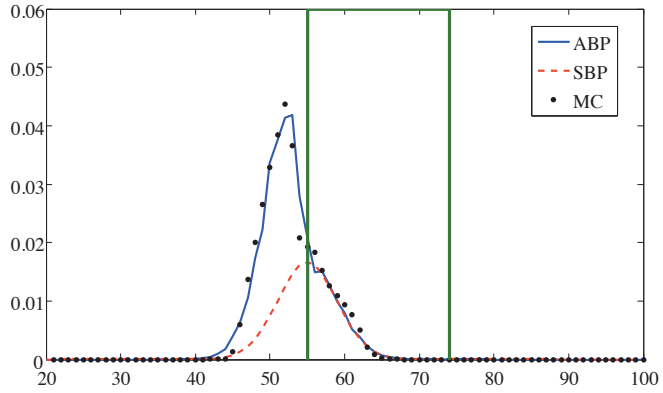


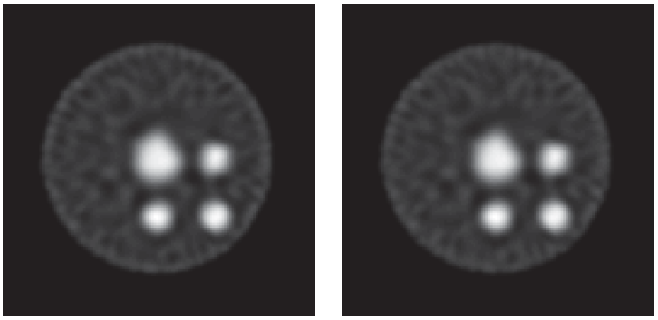
Fig. 4. Projected point source section using ABP and SBP projector and MC. The rectangle represents the border of the attenuation medium.

C. Monte-Carlo simulations

1) Point source in attenuated medium: We projected a simulated point-source located in a rectangular phantom containing water with the Monte-Carlo (MC) code SIMIND [16], as well as using SBP and ABP. The point-source was purposely located at the border of the attenuation medium in order to assess the effect of the central line approximation. A section of the projected point-source is shown in figure 4. It shows that the central line approximation results in an evenly distributed projected point-source using SBP, whereas the projected point source using ABP is similar to the MC projection.

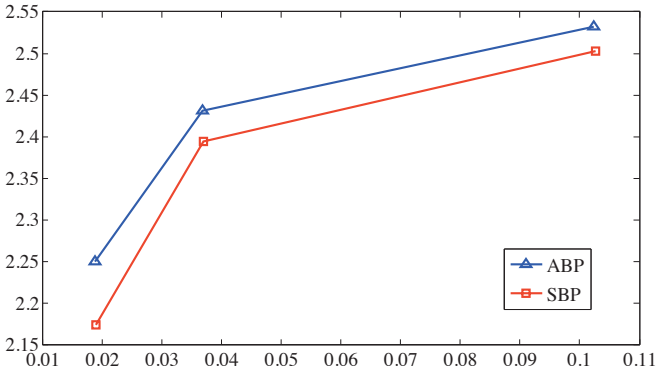
2) Phantom evaluation: We evaluated our new projector/back-projector using simulated data. MC SPECT projection data were generated using SIMIND. The activity distribution was a cylinder (28 cm diameter) containing 4 cylindrical inserts of different sizes (diameters from 35 to 56 mm). The true contrast in all spheres compared to the background was 3. Simulations were done corresponding to a rotating scintillation camera equipped with a LEHS collimator with a radius of rotation of $\rho = 192\,\text{mm}$. The number of projection angles over 360° was 120 ($n = 120 \times$ number of pixels/projection). The effects of scatter were not simulated. The object central slice was reconstructed in 2-D with ABP and SBP. ABP was performed using 720 azimuthal polar virtual angles ($N = 720 \times$ number of pixels/projection). Activity images were reconstructed with a surrogate based algorithm [7] with a quadratic smoothing prior weighted by a parameter β. We used 3 different values of the regularisation parameter β.

The reconstructed images were assessed by their mean contrasts in each cylinder as well as coefficient of variation (COV) calculated across 10 MC realisations. Reconstructed images using SBP and ABP are shown in figure 5(a) and 5(b) respectively. The 2 images appear similar although a weak dark ring can be seen in the SBP reconstructed image close to the edge of the phantom, probably due to inaccurate attenuation modelling. Figure 6 shows the COV vs contrast curves of the penalised maximum

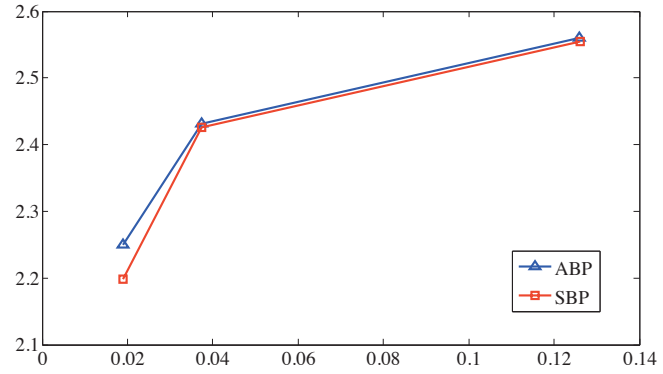


(a) SBP reconstruction (b) ABP reconstruction

Fig. 5. Reconstruction from MC data: (a) using SBP; (b) using ABP.



(a) 400 iterations



(b) 600 iterations

Fig. 6. Penalised-ML reconstruction contrast vs COV for 3 different values of β: (a) 400 iterations; (b) 600 iterations.

likelihood (ML) reconstructed images using SBP and ABP for 3 values of β after 400 iterations (figure 6(a)) and 600 iterations (figure 6(b)). The contrast was calculated over the bottom right disk. Although results are very similar after 600 iterations, ABP reconstruction performs better than SBP when only 400 iterations are performed.

IV. Discussion and conclusion

Here we have presented a new projection/back-projection technique for SPECT reconstruction, which is based on an angular blurring approach instead of the traditional distance dependent blurring approach. Our new method utilises the speed-up obtained with a GPU-device for parallel-beam forward and back-projection, and has a high degree of flexibility, allowing a wide range of collimators to be modelled by simply changing a weighting function. Here we have illustrated the flexibility of the method and we have shown that it produces results similar to the traditional approach. In further work we intend to model a wider range of imaging systems and optimise their performances.

Acknowledgment

The authors would like to thank Michael Ljungberg and Maria Holstensson for help with running SIMIND.

References

[1] G. T. Gullberg, R. H. Huesman, J. A. Malko, N. J. Pelc, and T. F. Budinger, "An attenuated projector-backprojector for iterative spectreconstruction," *Phys. Med. Biol.*, vol. 30, no. 8, pp. 799–816, 1985.

[2] G. L. Zeng, C. Bai, and G. T. Gullberg, "A projector/backprojector with slice-to-slice blurring for efficient three-dimensional scatter modeling," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 722–732, 1999.

[3] L. Lin, P. L. Kench, M.-C. Gregoire, and S. R. Meikle, "Projection process modelling for iterative reconstruction of pinhole SPECT," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 5, pp. 2578–2586, 2010.

[4] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, 1982.

[5] K. Lange and R. E. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomo.*, vol. 8, no. 2, pp. 306–316, 1984.

[6] P. J. Green, "Bayesian reconstruction from emission tomography using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–92, 1990.

[7] A. R. de Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–137, 1995.

[8] R. K. Rowe, J. N. Aarsvold, H. H. Barrett, J. C. Chen, W. P. Klein, B. A. Moore, I. W. Pang, D. D. Pattonand, and T. A. White, "A stationary hemispherical SPECT imager for three-dimensional brain imaging," *J. Nucl. Med.*, vol. 34, no. 3, pp. 474–480, 1993.

[9] F. van der Have, B. Vastenhouw, M. Rentmeester, and F. J. Beekman, "System calibration and statistical image reconstruction for ultra-high resolution stationary pinhole SPECT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 960–971, 2008.

[10] M. Rafecas, B. Mosler, M. Dietz, M. Pogl, A. Stamatakis, D. P. McElroy, and S. Ziegler, "Use of a Monte Carlo-based probability matrix," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2597–2605, 2004.

[11] D. Lazaro, Z. el Bitar, V. Breton, D. Hill, and I. Buvat, "Fully 3D Monte Carlo reconstruction in SPECT," *Phys. Med. Biol.*, vol. 2005, pp. 3739–3754, 2005.

[12] Z. el Bitar, Y. Petegnief, D. Lazaro, D. Hill, V. Breton, and I. Buvat, "Targeted fully 3D Monte Carlo reconstruction in SPECT," in *IEEE Nucl. Sci. Symp. Conf. Record,*, vol. 6, 2006, pp. 3414–3419.

[13] A. S. Fokas, B. F. Hutton, and K. Kacperski, "Modeling of collimator in analytical SPECT image reconstruction," internal report.

[14] S. Pedemonte, A. Bousse, K. Erlandsson, M. Modat, S. Arridge, B. F. Hutton, and S. Ourselin, "GPU accelerated rotation-based emission tomography reconstruction," in *IEEE Med. Imag. Conf. Record*, 2010, pp. 2657–2661.

[15] F. Natterer, *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, 2001.

[16] M. Ljungberg and S.-E. Strand, "A Monte Carlo program simulating scintillation camera imaging," *Comput. Methods Programs Biomed.*, vol. 29, pp. 257–272, 1989.

A novel method of embedding additional information into tensor diffusion filtering as an application for multi-modal reconstruction in ET

D. Kazantsev[1], S. Ourselin[2], B. F. Hutton[3], W.R.B. Lionheart[4] and S. R. Arridge[2]

***Abstract*—The fast development of hybrid imaging modalities in tomography, such as SPECT (single photon emission computed tomography)/CT (computed tomography), PET (positron emission tomography)/MRI (magnetic resonance imaging) and PET/CT, have increased an interest for reconstruction algorithms which are able to utilize a functional and anatomical information at the same time.**

In this paper a new method proposed for iterative reconstruction with anatomical prior in emission tomography (ET). The introduced regularization term is a modified anisotropic tensor diffusion filter which has shape-adapted smoothing properties. The filter accommodates available anatomical information which results in enhanced position and image dependent spatial resolution of emission images. Based on underlying orientations of normal and tangential vector fields for emission and anatomical images, the diffusion flux is rotated and scaled. Poisson likelihood fidelity and penalty terms are optimized separately by means of forward-backward splitting (FBS) technique. Presented approach is validated quantitatively using co-registered SPECT/MR synthetic data and compared with another anatomically penalized reconstruction as well as with iterative statistical reconstruction without regularization.

***Index Terms*—emission tomography, hybrid modalities, anatomical prior, image denoising, tensor diffusion, anisotropy, regularization, splitting methods**

I. Introduction

For reconstruction of activity distribution in ET, iterative statistical methods, such as, maximum-likelihood expectation maximization (MLEM) algorithm [1] are commonly used. Since the reconstruction problem in ET is ill-posed and ill-conditioned it requires additional regularization to ensure well-posedness. From a Bayesian perspective, imposing desirable properties (e.g. smoothness) on the solution leads to a prior probability characterization. A maximum *a posteriori* probability (MAP) estimate or penalized likelihood are successful methods employing the prior information. In the tomographic reconstruction, the sum of the likelihood penalized by the noise suppressing term is optimized [2].

The use of variational regularization techniques for tomographic reconstruction is significantly supported by the successful development of image processing tools and corresponding mathematical framework [3]. Successfully dealing with noise in images while leaving important features intact, the penalties based on partial differential equations (PDE), such as total variation (TV) [4] and anisotropic diffusion (AD) [5],[6] are competitive means to regularize reconstruction in emission tomography (ET).

The availability of side information from hybrid scanners (SPECT/CT, PET/MRI and PET/CT) can improve resolution of functional image by referring to prominent edges of anatomical data [7]. In this paper we propose a new method to smooth radiopharmaceutical distribution by means of available anatomical information. Although the proposed method has some similarities with previous work on modified diffusion filters [8], [9] it is based on a very new idea of embedding available data into the filtering process.

To reach a desired solution we utilize the FBS approach [10] for cost function iterative optimization. Cost function is decomposed into two sub-problems which are optimized separately. Similarly to [11] we split Poisson data fidelity and variational penalty terms. The resulting algorithm consists of the classical MLEM step followed by the modified diffusion step. The nature of the diffusion step is the main contribution of this paper.

II. Method

A. Tensor based anisotropic diffusion filtering (TBADF)

Following the Weickert's approach [5] for evaluation of local orientations of the image gradient $\nabla u(\mathbf{x})$ the *structure tensor* has to be build. Here we consider the 2D case ($\mathbf{x} = (x, y)$), however the proposed method can be easily generalized for the 3D. To avoid false edge detections due to noise, $u(\mathbf{x})$ is convolved with a Gaussian kernel k_σ, where σ is a *differentiation scale*: $u_\sigma(\mathbf{x}) = (k_\sigma * u)(\mathbf{x})$. The local information is averaged by convolving component-wise $\nabla u_\sigma \nabla u_\sigma^T$ with a Gaussian kernel k_ρ, where ρ is an *integration* scale which controls the size of the neighbourhood with gradients dominant orientations [3].

The structure tensor can be constructed as a symmetric, positive semidefinite (PSD) matrix:

$$\mathbf{J}_\rho(\nabla u_\sigma)(\mathbf{x}) = k_\rho * (\nabla u_\sigma \nabla u_\sigma^T). \quad (1)$$

Principal axis transformation of (1) gives the orthonormal eigenvectors $\boldsymbol{v_1} \parallel \nabla u_\sigma, \boldsymbol{v_2} \perp \nabla u_\sigma$, such as

$$\lim_{\rho\to 0} \boldsymbol{v_1} = \frac{\nabla u_\sigma}{|\nabla u_\sigma|} \quad ; \quad \lim_{\rho\to 0} \boldsymbol{v_2} = \frac{\nabla u_\sigma^\perp}{|\nabla u_\sigma|} \quad (2)$$

This work has been supported by the Engineering and Physical Sciences Research Council under grants EP/J010456/1, EP/I02249X/1 and EP/G026483/1. This research project has been supported by the European Commission under the 7th Framework Programme through the "Research Infrastructures" action of the "Capacities" Programme, NMI3-II Grant No. 283883, Project No.20120553

[1] The Manchester X-Ray Imaging Facility, School of Materials, The University of Manchester, Manchester, M13 9PL, UK (E-mail: daniil.kazantsev@manchester.ac.uk)

[2] Centre for Medical Image Computing, University College London, London NW1 9EE, UK

[3] Institute of Nuclear Medicine, UCL and University College London Hospitals NHS Trust, 235 Euston Road (T-5), London NW1 2BU, UK.

[4] School of Mathematics, Alan Turing Building, The University of Manchester, M13 9PL, UK

where $|\nabla u| = \sqrt{u_x^2 + u_y^2}$ and the corresponding eigenvalues defined using the PSD matrix components as:

$$\eta_{1,2} = \frac{1}{2}\left(j_{11} + j_{22} \pm \sqrt{(j_{11} - j_{22})^2 + 4j_{12}^2}\right), \quad (3)$$

The eigenvalues $\eta_{1,2}$ (averaged by the scale parameter ρ) convey the level of intensity propagation along the given directions $\boldsymbol{v_{1,2}}$. The eigenvalues characterize local geometrical information, for isotropic areas $\eta_1 \cong \eta_2 \cong 0$ and $\eta_1 \gg \eta_2$ or $\eta_1 \ll \eta_2$ for anisotropic (line structures).

Furthermore, based on $\eta_{1,2}$ one can estimate the level of anisotropy using the normalized measure of coherence [12]:

$$C(\mathbf{x}) = \left(\frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}\right)^2; \quad C(\mathbf{x}) \in [0, 1], \quad (4)$$

when $C(\mathbf{x}) = 1$ the gradient is totally aligned, when $C(\mathbf{x}) = 0$ it has no preferred direction. Note that the measure (4) is undefined for the uniform regions.

The diffusion PDE in its general form can be written as:

$$\begin{cases} \partial_t u = \nabla \cdot (\mathbf{D}(\mathbf{J}_\rho(\nabla u_\sigma))\nabla u), \\ u(\mathbf{x}, 0) = u_0. \end{cases} \quad (5)$$

For $\mathbf{D} = \mathbf{I}$, (5) performs a linear isotropic diffusion, for $\mathbf{D} = \varphi(|\nabla u|)\mathbf{I}$ it is equivalent to a nonlinear isotropic diffusion (φ is an edge preserving function [5]), and for

$$\mathbf{D} = [\boldsymbol{v_1}\ \boldsymbol{v_2}] \cdot \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{v_1}^T \\ \boldsymbol{v_2}^T \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix}, \quad (6)$$

equation (5) stands for anisotropic nonlinear diffusion driven by the diffusion tensor with elements:

$$D_{i,j} = \sum_{n=1,\ldots,2} \gamma_n v_{ni} v_{nj}, \quad (7)$$

The diffusion tensor (6) has a new eigenvalues $\gamma_{1,2}$ which define the strength of smoothing in the preferred directions $\boldsymbol{v_{1,2}}$. A few different definitions for γ_n related to the edge enhancing diffusion (EED) and the coherence enhancing diffusion (CED) were proposed by Weickert [13]. In this paper we will be using the EED approach, where smoothing in normal direction is reduced by:

$$\gamma_1 = \varphi(|\nabla u|) = \exp\left(-|\nabla u|^2/\epsilon^2\right), \quad (8)$$

while the eigenvalue related to the tangential vector is $\gamma_2 = 1$. The threshold parameter ϵ controls the strength of diffusion. The diffusion tensor rotates and scales the flux in order to adapt it for underlying geometrical configurations of image $u(\mathrm{x})$.

B. Embedding additional information into the TBADF

Here we present a novel idea how to embed an additional information $\boldsymbol{\mu}$ into the image $\boldsymbol{\lambda}$ by means of the tensor based diffusion filtering.

Let us assume that the diffusion tensors $\mathbf{D}_\mu$ and $\mathbf{D}_\lambda$ are defined for images $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ respectively.

The *combined* diffusion tensor can be defined using following arithmetic interpolation scheme between two given tensors [14] as:

$$\mathbf{D}_{\mu,\lambda} = s\mathbf{D}_\lambda + (1 - s)\mathbf{D}_\mu; \quad s(\mathbf{x}) \in [0, 1], \quad (9)$$

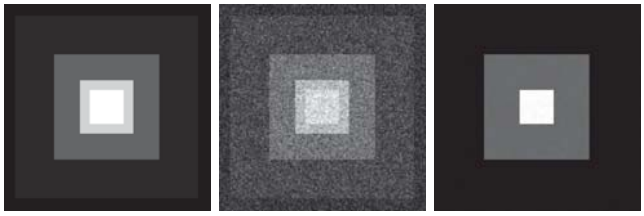


Fig. 1: From left to right: Original image, noisy image $\boldsymbol{\lambda}$ containing 10% of random noise and reference image $\boldsymbol{\mu}$ containing 1% of random noise.

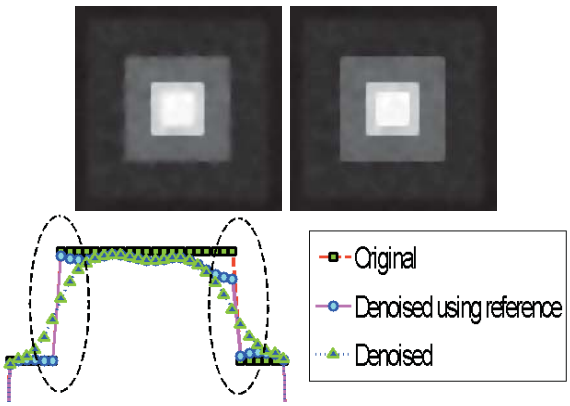


Fig. 2: Example of the TBAD filtering with a help of additional information. Top row from left to right: denoised image $\boldsymbol{\lambda}$ without use of the referenced image $\boldsymbol{\mu}$; filtered $\boldsymbol{\lambda}$ using the referenced image $\boldsymbol{\mu}$ and the combined tensor (9). Bottom picture is a plot of the horizontal middle section (the central region) of the denoised images above and the original image. In the marked areas edges are better preserved for the proposed method.

where $s(\mathrm{x})$ is a spatially variant parameter which should fulfil the following properties:

Property (1): $\forall \mathbf{x}$ calculate $\theta(\mathbf{x}) = acos(\boldsymbol{v_1}(\boldsymbol{\mu}) \cdot \boldsymbol{v_1}(\boldsymbol{\lambda}))$; if $\theta(\mathbf{x}) \in [0°, 0° + \theta_a]$ or $\theta \in [180° - \theta_a, 180°]$, then $s(\mathbf{x}) = 0$.

Property (2): $\forall s(\mathbf{x}) \neq 0$: $s(\mathbf{x}) = C_\lambda(\mathbf{x})$ (4).

Property (3): $\forall \mathbf{x}$ where $C_\lambda(\mathbf{x})$ is undefined: $s(\mathbf{x}) = 0$.

The ideas behind properties (1-3) are the following:

Property (1) is fulfilled when $C_\lambda(\mathbf{x}) \cong C_\mu(\mathbf{x}) \gg 0$, which is the case of an expressed anisotropy for both images. By checking angles between the two principal vector bases for images $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ one can estimate how bases are agreed with each other. An *acceptance angle* θ_a is introduced as a threshold parameter to make a decision on overlapping degree of two vectors. Empirically we found that $\theta_a = 10°$ gives satisfying results. When property (1) is fulfilled for some $\mathbf{x}$, so $s(\mathbf{x}) = 0$, it gives $\mathbf{D}_{\mu,\lambda} = \mathbf{D}_\mu$, then the diffusion (5) is performed using diffusion tensor of image $\boldsymbol{\mu}$.

Property (2) ensures that all values of $s(\mathbf{x})$ which are not equal to zero should be equal to $C_\lambda(\mathbf{x})$. Based on the level of anisotropy of image $\boldsymbol{\lambda}$ the interpolation between two tensors (9) takes place. For areas with higher anisotropy on $\boldsymbol{\lambda}$ there is less influence of $\boldsymbol{\mu}$.

Property (3) considers the uniform regions of image $\boldsymbol{\lambda}$ where C_λ is undefined ($\eta_1(\lambda) + \eta_2(\lambda) = 0$). If C_μ is undefined as well, then $\mathbf{D} = \mathbf{I}$.

The proposed method was tested for the simple denoising

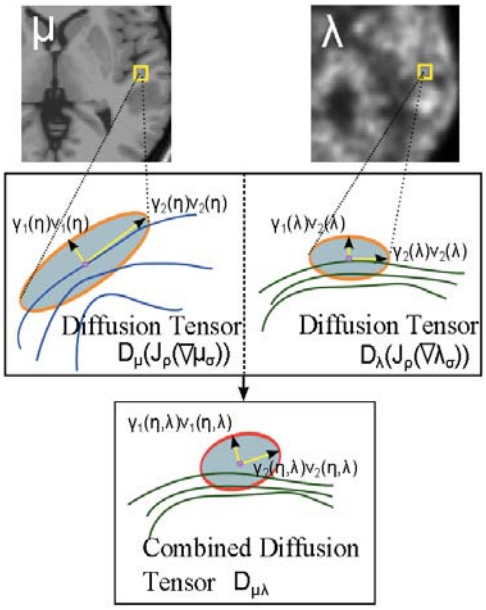


Fig. 3: The example of local geometrical structure represented by two diffusion tensors for anatomical $\boldsymbol{\mu}$ and emission images $\boldsymbol{\lambda}$. The combined diffusion tensor is constructed by interpolation (9).

procedure of image $\boldsymbol{\lambda}$ having additional image $\boldsymbol{\mu}$ as a reference, see Fig. 1. In Fig. 2 the results of the filtered image $\boldsymbol{\lambda}$ without use of referenced image $\boldsymbol{\mu}$ and with it. Note that for places where edges of exact image matched the reference sharp features are preserved, see plot in Fig. 2. This is the case when the property (1) is taken place. For areas where there is no additional information from $\boldsymbol{\mu}$ the property (2) performs.

The idea of the combined tensor filtering (9) can be used for the hybrid imaging modalities having emission data registered with anatomical one. The Fig. 3 explains how the combined tensor can be used for the problem of reconstruction in ET having data from MR scanner.

C. An anatomically driven tensor based anisotropic diffusion filtering (ADTBADF) for ET reconstruction

Lets consider the discrete activity distribution as N-dim vector $\boldsymbol{\lambda}$ with elements $\lambda_i, i = 1, \ldots, \mathrm{N}$ and $\boldsymbol{g}$ is a measured projection data (sinogram) with elements $g_j, j = 1, \ldots, \mathrm{M}$. In the ET, $\boldsymbol{g}$ follows Poisson distribution with mean $\overline{\boldsymbol{g}} = \mathrm{P}\boldsymbol{\lambda}$, where projection or system matrix $\mathrm{P} : \mathbb{R}^N \to \mathbb{R}^M$ depends on system design and detector array geometry.

To reconstruct desired image $\boldsymbol{\lambda}$ from Poisson distributed data $\boldsymbol{g}$, the following constrained cost function needs to be optimized:

$$\min_{\boldsymbol{\lambda}} \left\{ \mathbb{D}_{KL}(\boldsymbol{g}, \mathrm{P}\boldsymbol{\lambda}) + \beta \int_{\Omega} h(|\nabla \boldsymbol{\lambda}|) \mathrm{d}\Omega \right\} \quad \text{s.t. } \boldsymbol{\lambda} \geq 0, \quad (10)$$

where $\mathbb{D}_{KL}(\boldsymbol{g}, \mathrm{P}\boldsymbol{\lambda}) = \int [\overline{\boldsymbol{g}} - \boldsymbol{g} \log \overline{\boldsymbol{g}}]$ is a Kullback - Leibler distance and second term is a convex energy functional controlled by regularization parameter β [15].

The algorithms based on proximity operator properties [10] can solve (10) by splitting regularization and data-fidelity terms in a way that two (generally less complex) sub-problems have to be solved.

Similarly to [11], [8] a nested two step iteration algorithm can be derived in a form:

$$\begin{array}{rcll} \boldsymbol{\lambda}^{m+\frac{1}{2}} & = & \frac{\boldsymbol{\lambda}^m}{\mathrm{P}^*\mathbf{1}} \mathrm{P}^* \left(\frac{\boldsymbol{g}}{\mathrm{P}\boldsymbol{\lambda}^m} \right), & \textit{MLEM step} \\ \boldsymbol{\lambda}^{m+1} & = & \mathrm{L}\left(\boldsymbol{\lambda}^{m+\frac{1}{2}} \right) & \textit{Filtering step} \end{array} \quad (11)$$

Where L is a diffusion operator which performs transition from $\boldsymbol{\lambda}^{m+\frac{1}{2}}$ to $\boldsymbol{\lambda}^{m+1}$ by minimizing the following function:

$$\Psi(\boldsymbol{\lambda}) = \frac{1}{2} \int_{\Omega} \frac{\mathrm{P}^*\mathbf{1}}{\boldsymbol{\lambda}^m} \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{m+\frac{1}{2}} \right)^2 + \beta \int_{\Omega} h(|\nabla \boldsymbol{\lambda}|) \mathrm{d}\Omega \quad (12)$$

In this work we use a standard iterative gradient descent algorithm to optimize (12):

$$\begin{cases} \boldsymbol{\lambda}^{l+1} = \boldsymbol{\lambda}^l + \tau(\Psi'(\boldsymbol{\lambda})) \\ \boldsymbol{\lambda}^1 = \boldsymbol{\lambda}^{m+\frac{1}{2}}, \end{cases} \quad (13)$$

$$\Psi'(\boldsymbol{\lambda}) = \frac{\mathrm{P}^*\mathbf{1}}{\boldsymbol{\lambda}^m} \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}^{m+\frac{1}{2}} \right) + \beta(\nabla \cdot (\mathbf{D}_{\mu,\lambda} \nabla \boldsymbol{\lambda})). \quad (14)$$

The right part of (14) includes the sum of weighted data fidelity term and a nonlinear anisotropic diffusion term which includes the proposed combined tensor (9).

III. Numerical results

In this work we compare several reconstruction techniques, namely: an anatomically driven tensor based anisotropic diffusion filtering (ADTBADF) (11) with a filtering step (14), an anatomically driven nonlinear isotropic diffusion filtering (ADNIDF) with quadratic function (QADNIDF), with Huber function (HADNIDF) and classical MLEM (step 1 in (11)). The ADNIDF method is an iterative application of a Bowsher prior (a local activity smoothing technique based on intensity distribution in the neighbourhood of anatomical image, more in [16]) in the second step of (11). More information about this method can be found in [8].

For our experiments we used 2D synthetic Brainweb data to perform quantitative analysis of methods. The proposed algorithm can be generalized for 3D data, however the quantitative validation is much more time consuming in this case. The synthetic activity phantom is projected to form noise-free sinogram, then 30 Poisson noise realizations were generated from the data. Due to lack of space we are not presenting all necessary graphs for bias-variance analysis but we will comment on them.

In Fig. 4 one can see that the MLEM image has poor resolution and the highest variance for the hot lesion region of interest (ROI). However, the level of intensity in lesion ROI is high and the bias is very low. The use of quadratic smoothing with QADNIDF gives a significant enhancement of resolution of emission image (the smallest variance). The bias of QADNIDF is the highest due to over-smoothing of activity. The use of Huber function in HADNIDF can increase bias but as well as the variance. The reconstructed image with HADNIDF is less smooth (piecewise-constant appearance due to penalty function), but lesion is better quantified than with QADNIDF. The proposed approach (ADTBADF) allows to reduce bias further for approximately the same level of

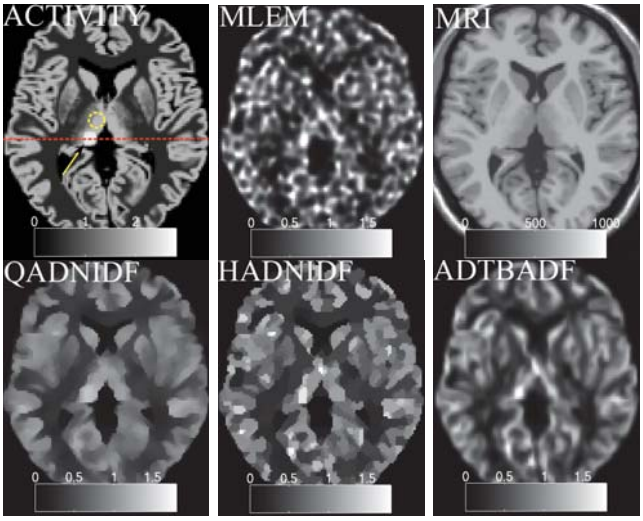


Fig. 4: left to right, top row: the 2D modeled activity phantom with hot lesion pointed by the arrow (the dotted circle was taken as a background region to calculate contrast for the lesion nearby and the background variance, MLEM reconstruction ($m = 170$) co-registered MR image; bottom row: reconstruction with QADNIDF, HADNIDF and ADTBADF.

variance. The value of variance is lowest for the gray matter region but the bias is slightly higher than for ADNIDF (edges are blurred due to convolution).

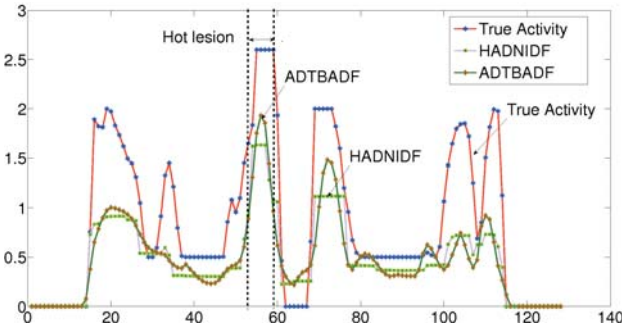


Fig. 5: 1D slices of the the true activity (horizontal line across the phantom in the Fig. 4), HADNIDF and ADTBADF images.

In the Fig. 5 one can see 1D slices of reconstructed images comparatively to original phantom. The proposed method gives smoother reconstruction but preserves an important features in data.

IV. DISCUSSION

The proposed algorithm has four additional parameters to control reconstruction comparatively to ADNIDF method (two Gaussian kernels σ, ρ in (1), the edge preserving threshold ϵ (8) for image μ and the angle of acceptance θ_a. However it doesn't add any significant difficulties to choose these parameters, first three can be estimated automatically based on the level of noise in image μ. The optimal threshold parameter ϵ is usually harder to determine in blurred images, a few suggestions for selection exist in literature [5]. In this work we didn't concentrate on this issue and the choice was empirical.

Here we present a binary decision procedure for the property (1) based on θ_a threshold. However due to noise and blur the local estimation of angles can be erroneous. Considering principal directions in a non-local region can improve property (1).

In ET reconstructed images are generally blurred (premature stop of the MLEM algorithm or smoothing penalty applied). The smooth appearance of the reconstructed images with the proposed method can be beneficial for clinicians who get used to visual assessment of similar images. The strength of blur for reconstructed images is controlled by width of Gaussian kernels and can be optimized.

V. CONCLUSION

In this work we present a novel approach of incorporating available additional information into diffusion filtering. The process of embedding information is performed by scaling and rotating the combined diffusion tensor of two images. Special properties of the proposed tensor were tested on the simple denoising example. The method shows the ability to detect matching directions of tensor fields of two images, it helps to retain edges and resolution in a filtered image.

The proposed method is tested for the modelled case of the SPECT reconstruction with co-registered MR data as a reference image. It produces better results in terms of bias, contrast and variance for lesions. Reconstructed images look more favourable due to smooth appearance and well preserved important features such as lesions.

REFERENCES

[1] A J Rockmore and A Macovski, "A maximum likelihood approach to emission image reconstruction from projections," *IEEE Trans. Nucl. Sci.*, vol. 23, pp. 1428–1432, 1976.
[2] Alvaro R De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, pp. 132–137, 1995.
[3] G Aubert and P Kornprobst, "Mathematical problems in image processing. partial differential equations and the calculus of variations," *Applied Mathematical Sciences.Springer.New York*, vol. 147, 2001.
[4] L I Rudin, S Osher, and E Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.
[5] J Weickert, "Anisotropic diffusion in image processing," *Stuttgart, Germany: Teubner*, 1998.
[6] P Perona and J Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis*, vol. 12, pp. 629–639, 1990.
[7] K Erlandsson, I Buvat, P H Pretorius, B A Thomas, and Hutton B F, "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology," *Physics in Medicine and Biology*, vol. 57, no. 21, pp. R119, 2012.
[8] D Kazantsev, S R Arridge, S Pedemonte, A Bousse, K Erlandsson, B F Hutton, and S Ourselin, "An anatomically driven anisotropic diffusion filtering method for 3d spect reconstruction," *Phys Med. Biol.*, vol. 57, pp. 3793–3810, 2012.
[9] C Chan, R R Fulton, D D Feng, and S R Meikle, "Regularized image reconstruction with an anatomically adaptive prior for positron emission tomography," *Phys Med. Biol.*, vol. 54, no. 24, pp. 7379–7400, 2009.
[10] P L Combettes and J C Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Eds. New York: Springer-Verlag*, vol. 10, 2010.
[11] A Sawatzky, C Brune, F Wubbeling, T Kosters, K Schafers, and M Burger, "Accurate em-tv algorithm in pet with low snr," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, 2008, pp. 5133–5137.
[12] B Jahne, "Spatio-temporal image processing: Theory and scientific applications," *Berlin: Springer-Verlag*, vol. 751, no. 0, 1993.
[13] J Weickert, "Coherence enhancing diffusion," *International Journal of Computer Vision*, vol. 31, pp. 111–127, 1999.
[14] P G Batchelor, M Moakher, Atkinsonm D, F Calamante, and A Connelly, "A rigorous framework for diffusion tensor calculus," *Magn. Reson. Med.*, vol. 53, pp. 221–225, 2005.
[15] J Qi and R M Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys Med. Biol.*, vol. 51, pp. R541–R578, 2006.
[16] K Vunckx, A Atre, K Baete, A Reilhac, CM Deroose, K Van Laere, and J Nuyts, "Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging," *IEEE Trans. Med. Imag.*, vol. 31(3), pp. 599–612, 2012.

Imaging Quality Improvement and Dose Reduction on TomoTherapy via Tensor Framelet: Phantom Study

Hao Gao, X. Sharon Qi, and Daniel A. Low

***Abstract*—This work is to investigate the feasibility of the megavoltage imaging quality improvement and imaging dose reduction for TomoTherapy, based on tensor framelet. That is, we aim to develop an effective and rapid reconstruction technique for the CT imaging on TomoTherapy to improve the imaging quality from the filtered backprojection, with full projection views or undersampled projection views. The phantom studies suggest that the tensor framelet method is robust for the low-dose imaging on TomoTherapy with 75% reduction of the projection views.**

***Index Terms*—TomoTherapy, low-dose, CT, tensor framelet, GPU, reconstruction.**

I. Introduction

INTENSITY Modulated Radiation Therapy (IMRT), capable of delivering highly conformal dose to the tumor while sparing the adjacent normal structures, is becoming the standard treatment for head-and-neck (H&N) and prostate cancer since the late 90s [1, 2]. However, rapid dose falloff of IMRT plans generally calls for more reproducible patient positioning and therefore accurate treatment delivery.

TomoTherapy Hi-Art Helical Radiotherapy System (Accuray, Sunnyvale, CA) is an integrated unit dedicated to IMRT using megavoltage CT (MVCT) as image guidance [3-5]. The utility of on-board CT detector ensures the accuracy of patient's anatomy and alignment using a relatively low-dose on daily basis, but also allows us to assess treatment delivery verification for each radiation treatment by collecting MVCT data immediate prior or after treatment delivery. Large difference between the delivered dose and the planned dose generally implies an adaptive therapy such as re-planning.

However, concerns have been raised for additional doses to patients due to daily image scans, which might lead to unnecessary toxicity for sensitive structures [6-8]. It has been a great challenge to improve the MVCT image quality while maintain low-dose due to imaging [9].

On the other hand, inspired by compressive sensing [10, 11], there have been many efforts on iterative reconstruction methods for the low-dose image reconstruction with undersampled CT or dynamic CT data, using the L1-type image regularization [12-18]. In our recent work on 4D cone beam CT [19], we have proposed the tensor framelet (TF), which is ideal for high-dimensional large-scale image reconstruction due to its significantly reduced demand on the memory and computational cost.

In this work, using TF, we aim to develop new reconstruction method to further improve the MVCT imaging quality on TomoTherapy. Moreover, we investigate the possibility of the low-dose imaging with fewer projection views, to produce the comparable imaging quality.

This work is partially supported by NIH/NIBIB grant EB013387.

H. Gao is with the Departments of Mathematics and Computer Science, and Radiology and Imaging Sciences, Emory University, Atlanta, GA 30322 USA (e-mail: hao.gao@emory.edu).

X. S. Qi and D. A. Low are with the Department of Radiation Oncology, University of California, Los Angeles, CA 90095, USA.

II. Methodology

A. Least-square Formulation

With the traditional FBP, the 3D CT images on TomoTherapy could be reconstructed slice by slice along the longitudinal direction based on the fan-beam geometry with curved detectors. To utilize the prior that the CT image or its derivates can be smooth and sparse for both the in-plane directions and the longitudinal direction, we formulate the image reconstruction as the following iterative least-square minimization problem, in which all slices are reconstructed simultaneously so that the image smoothness and sparsity along the longitudinal direction can be enforced,

$$X = \arg\min_X \frac{1}{2} \| AX - Y \|_2^2 + \lambda R(X). \quad (1)$$

In (1), the first term is the L2-norm data fidelity term with the imaging data Y and the 3D image X to be reconstructed, and the second term is the L1-norm image regularization term with the regularization parameter λ and the proper sparsifying transform, which will be discussed next.

Here A is a linear operator on X that corresponds to the X-ray transform on X slice by slice. Considering the computational efficiency, we use our recently developed new parallel algorithm with O(1) per parallel thread [20, 21].

In this paper, we consider two reconstruction settings with coarse and fine resolution respectively along the longitudinal direction. For the coarse-resolution reconstruction, every 360-degree projection views are used to reconstruct one slice; for the fine-resolution reconstruction, every 360-degree projection views are used to reconstruct two slices, i.e., 180-degree projection views for one slice.

B. Tensor Framelet

We formulate TF [19, 22] at multilevel based on the 1D framelet operator *w*. Considering the 1D framelet transform up to *L* levels (larger number for coarser resolution), we start from the 1D refinement masks for *w* at Level $0 \le l \le L$

$$h^l = \begin{bmatrix} h_0^l \\ h_1^l \\ h_2^l \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & \underbrace{0 \cdots 0}_{2^l-1} & 2 & \underbrace{0 \cdots 0}_{2^l-1} & 1 \\ \sqrt{2} & \underbrace{0 \cdots 0}_{2^l-1} & 0 & \underbrace{0 \cdots 0}_{2^l-1} & -\sqrt{2} \\ -1 & \underbrace{0 \cdots 0}_{2^l-1} & 2 & \underbrace{0 \cdots 0}_{2^l-1} & -1 \end{bmatrix}. \quad (2)$$

Then 1D framelet transform *w* of *x* is

$$wx = \left[\underbrace{d_1^0 x^0 \quad d_2^0 x^0}_{\text{Level: } 0} \cdots \underbrace{d_1^l x^l \quad d_2^l x^l}_{\text{Level: } l} \cdots \underbrace{x^{L+1} \quad d_1^L x^L \quad d_2^L x^L}_{\text{Level: } L} \right], \quad (3)$$

and its transpose w^T is

$$w^T(wx) = \sum_{m=1}^{2} \sum_{l=0}^{L} d_m^l x^l + x^{L+1}, \quad (4)$$

where

$$d_m^l x^l = h_m^l * x^l, \quad x^{l+1} = d_0^l x^l \quad (5)$$

with * for convolution, $x^0 = x$, $0 \le m \le 2$, and $0 \le l \le L$.

Based on (3) and (4), the TF with multilevel is

$$WX = \frac{1}{\sqrt{3}} \begin{bmatrix} w_x X_x \\ w_y X_y \\ w_z X_z \end{bmatrix}, \quad (6)$$

and the adjoint TF with multilevel is

$$W^T(WX) = \frac{1}{3}[\sum_{j,k} w_x^T(w_x X_x^{jk}) + \sum_{i,k} w_y^T(w_y X_y^{ik}) + \sum_{i,j} w_z^T(w_z X_z^{ij})], \quad (7)$$

where X_x, X_y, X_z are the unfolded matrices of X along x, y, z dimension respectively, and the 1D framelet operator w and w^T are with respect to the 1D unfolded dimension x, y, z respectively. For example, $w_x X_x$ performs 1D framelet transform along each x-line for all combination of y- and z-variables.

Finally, the isotropic TF norm at multilevel is defined as

$$\| WX \|_1 = \frac{1}{\sqrt{3}} (\sum_{m=1}^{2} \sum_{l=0}^{L} \lambda_m^l \| D_m^l X^l \|_1 + \lambda_0^{L+1} \| X^{L+1} \|_1), \quad (8)$$

with

$$\| D_m^l X^l \|_1 = \sum_{i,j,k} \sqrt{|D_{mx}^l X_x^l|^2 + |D_{my}^l X_y^l|^2 + |D_{mz}^l X_z^l|^2}, m = 1, 2 \quad (9)$$

and

$$\| X^{L+1} \|_1 = \sum_{i,j,k} \sqrt{|X_x^{L+1}|^2 + |X_y^{L+1}|^2 + |X_z^{L+1}|^2}. \quad (10)$$

Notice that the wavelet bases are orthonormal, while the TF bases are redundant. For example, the Haar wavelet includes the low-passed average and the first-order derivatives, while the piecewise-linear TF here also contains the second-order derivatives. In this sense, TF generalizes the wavelet with high-order derivatives for characterizing the smoothness and the sparsity.

With TF, the formulation is a L1-norm-regularized least-square optimization. Here we choose the Split Bregman method [23] for solving this convex L1-type problem. The method was also used in our prior related work on CT [17,19,24].

III. Materials

We scanned a H&N anthropomorphic phantom, and a pelvis anthropomorphic phantom on our TomoTherapy HD unit using normal MVCT scan thickness. The detector data were exported out for each phantom immediately after each MVCT scan, an air scan was also acquired to normalize the raw detector output. Default image scanning parameters (TomoTherapy V4.2) were used in this study: jaws were set to a position of ±0.5 mm (J1), and the projected beam width at isocenter was approximately 4mm, gantry period of 10 seconds with couch speed of 8mm/rotation. The repetition rate for imaging mode was 80 Hz.

As the geometric parameters for image reconstruction, the distance from source to the isocenter was 85 cm, the distance from the source to the detector was 144 cm. The detector array had 640 pixels and the central element was offset by 29.5 pixels. There were 800 projection views per rotation and the couch speed was 8 mm per rotation. For the current system, the pixels from the 27th to the 554th were available for image reconstruction. Since the center of the actual curved detector array was different from the isocenter for TomoTherapy system to improve the efficiency of the outer channels, the virtual curved detector array centered at the isocenter was created with 0.048 degrees as the angular pixel size. As a result, the reconstructed image per slice was 350×350 with 0.1 cm×0.1 cm. The same imaging parameters were used for TF and FBP.

The GPU-based reconstruction was implemented with a NVIDIA GeForce GTX 680 GPU card (1536 cores and 2.0 GB device memory). it took ~2 minutes for our GPU-based solver to reconstruct a 80-slice 350×350 3D image with 100 projection views per slice and 528 detections per view.

To evaluate the imaging quality without undersampling, we first performed the coarse reconstruction (360-degree projection views per slice) with 800 projection per slice, and then the fine reconstruction (180-degree projection views per slice) with 400 projection per slice.

To evaluate the imaging quality with undersampling, we first performed the coarse reconstruction (360-degree projection views per slice) with 200 projection per slice (25% data) and 100 projection per slice (12.5% data) respectively, and then the fine reconstruction (180-degree projection views per slice) with 100 projection per slice (25% data) and 50 projection per slice (12.5% data) respectively.

IV. Results

A. H&N Phantom Study

A coarse-resolution slice (360-degree projection views per slice), and the corresponding two fine-resolution slices (180-degree projection views per slice) are plotted in Fig. 1, 2, and 3 respectively. That is the total amount of 180-degree data for reconstructing two slices in Fig. 2 and 3 is the same as the amount of 360-degree data for reconstructing the single slice in Fig. 1. Here the coarse and the fine are with respect to the resolution along the longitudinal direction.

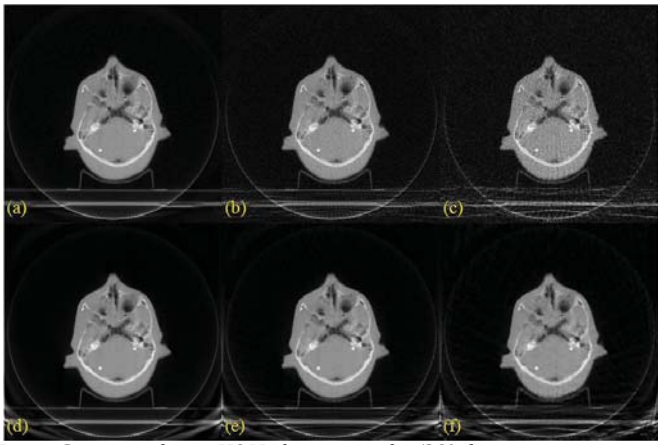


Fig. 1. Coarse-resolution H&N phantom results (360-degree projections per slice). (a), (b), and (c) are from FBP with 800 projections (100%), 200 projections (25%), and 100 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 800 projections (100%), 200 projections (25%), and 100 projections (12.5%) per slice respectively. Here the 360-degree projection views are the same as the combined 180-degree projection views for Fig. 2 and 3.

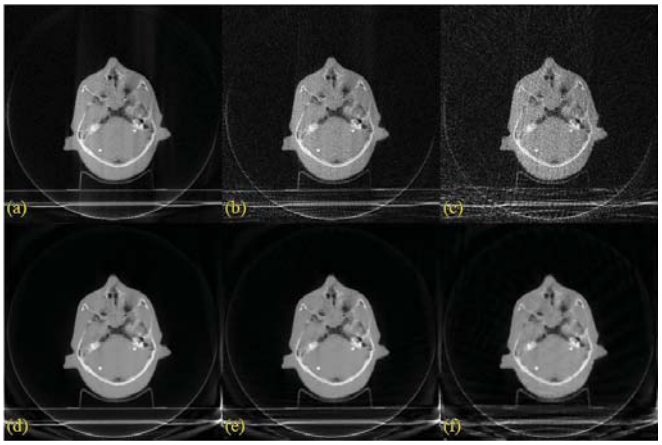


Fig. 2. Fine-resolution H&N phantom results (180-degree projections per slice). (a), (b), and (c) are from FBP with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively. Here the 180-degree projection views together with those for Fig. 3 are the same as the 360-degree projection views for Fig. 1.

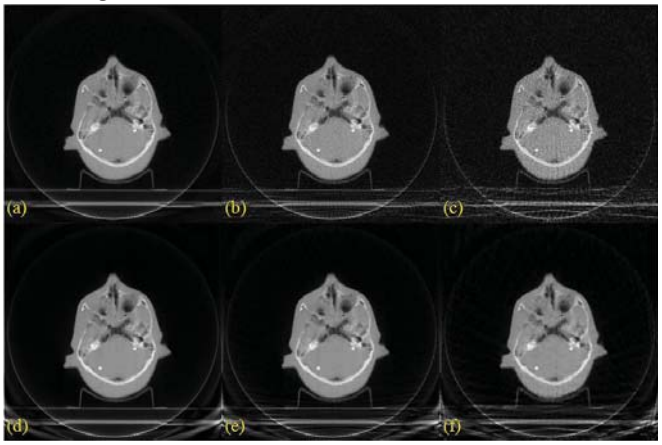


Fig. 3. Fine-resolution H&N phantom results (180-degree projections per slice). (a), (b), and (c) are from FBP with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively. Here the 180-degree projection views together with those for Fig. 2 are the same as the 360-degree projection views for Fig. 1.

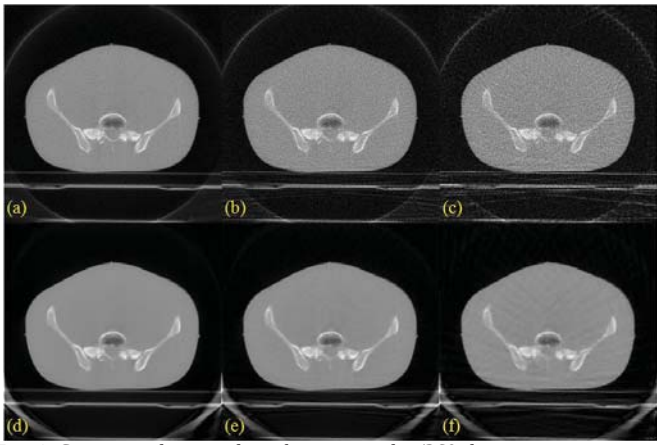


Fig. 4. Coarse-resolution pelvis phantom results (360-degree projections per slice). (a), (b), and (c) are from FBP with 800 projections (100%), 200 projections (25%), and 100 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 800 projections (100%), 200 projections (25%), and 100 projections (12.5%) per slice respectively. Here the 360-degree projection views are the same as the combined 180-degree projection views for Fig. 5 and 6.

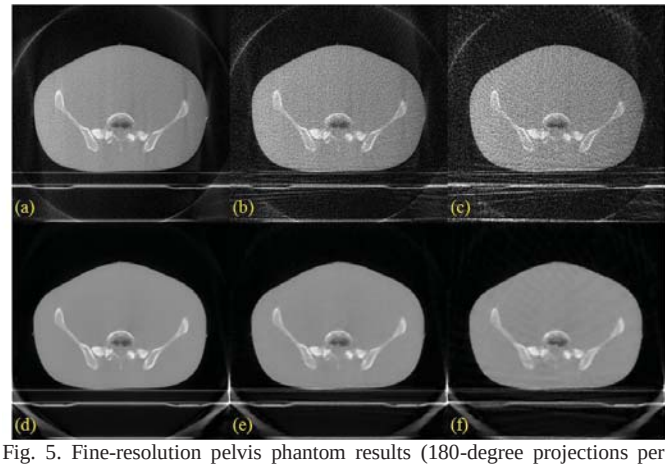


Fig. 5. Fine-resolution pelvis phantom results (180-degree projections per slice). (a), (b), and (c) are from FBP with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively. Here the 180-degree projection views together with those for Fig. 6 are the same as the 360-degree projection views for Fig. 4.

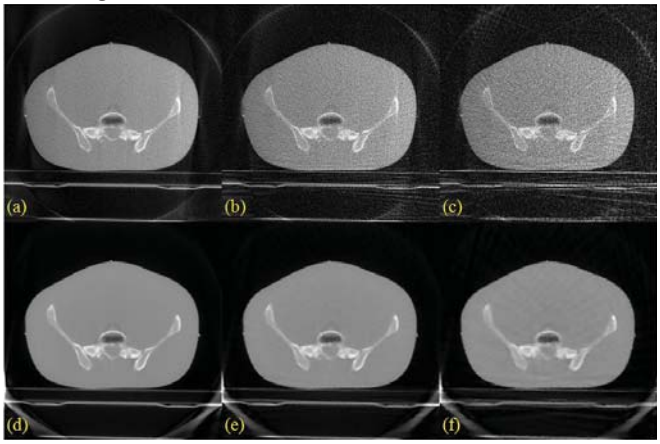


Fig. 6. Fine-resolution pelvis phantom results (180-degree projections per slice). (a), (b), and (c) are from FBP with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively; (c), (d), and (f) are from TF with 400 projections (100%), 100 projections (25%), and 50 projections (12.5%) per slice respectively. Here the 180-degree projection views together with those for Fig. 5 are the same as the 360-degree projection views for Fig. 4.

With 100% data, it is clear that TF (i.e., Fig. 1(d), 2(d), 3(d)) provides apparently better image quality than FBP (i.e., Fig. 1(a), 2(a), 3(a)) with less artifacts.

With 25% data, TF (i.e., Fig. 1(e), 2(e), 3(e)) still provides apparently better image quality than FBP (i.e., Fig. 1(b), 2(b), 3(b)). Moreover, the image quality via TF with 25% data (i.e., Fig. 1(e), 2(e), 3(e)) is comparable to the image quality via TF with 100% data (i.e., Fig. 1(d), 2(d), 3(d)), and the image difference, i.e., $||X\text{-}X_0||_1/||X_0||_1$, between X (from TF with 25% data) and X_0 (from TF with 100% data) within the central ROI of the 256-pixel diameter is 2.8%, 3.2%, 3.4% for Fig. 1, 2, 3 respectively.

With 12.5% data, the reconstructed image via TF (i.e., Fig. 1(f), 2(f), 3(f)) remains significantly better than that from FBP (i.e., Fig. 1(c), 2(c), 3(c)). And the image difference between X (from TF with 12.5% data) and X_0 (from TF with 100% data) within the central ROI of the 256-pixel diameter is 4.7%, 5.5%, 5.9% for Fig. 1, 2, 3 respectively.

B. Pelvis Phantom Study

A coarse-resolution slice (360-degree projection views per slice), and the corresponding two fine-resolution slices (180-degree projection views per slice) are plotted in Fig. 4, 5, and 6 respectively. That is the total amount of 180-degree data for reconstructing two slices in Fig. 5 and 6 is the same as the amount of 360-degree data for reconstructing the single slice in Fig. 4. Here the coarse and the fine are with respect to the resolution along the longitudinal direction.

With 100% data, it is clear that TF (i.e., Fig. 4(d), 5(d), 6(d)) provides apparently better image quality than FBP (i.e., Fig. 4(a), 5(a), 6(a)) with less artifacts.

With 25% data, TF (i.e., Fig. 4(e), 5(e), 6(e)) still provides apparently better image quality than FBP (i.e., Fig. 4(b), 5(b), 6(b)). Moreover, the image quality via TF with 25% data (i.e., Fig. 4(e), 5(e), 6(e)) is comparable to the image quality via TF with 100% data (i.e., Fig. 4(d), 5(d), 6(d)), and the image difference, i.e., $||X\text{-}X_0||_1/||X_0||_1$, between X (from TF with 25% data) and X_0 (from TF with 100% data) within the central ROI of the 300-pixel diameter is 1.8%, 2.3%, 2.2% for Fig. 4, 5, 6 respectively.

With 12.5% data, the reconstructed image via TF (i.e., Fig. 4(f), 5(f), 6(f)) remains significantly better than that from FBP (i.e., Fig. 4(c), 5(c), 6(c)). And the image difference between X (from TF with 12.5% data) and X_0 (from TF with 100% data) within the central ROI of the 300-pixel diameter is 3.2%, 3.8%, 3.9% for Fig. 4, 5, 6 respectively.

V. CONCLUSOINS

We have proposed a novel TF-based image reconstruction technique that provides better image quality than FBP for the megavoltage CT imaging on TomoTherapy with full or undersampled projection views. In particular, the phantom studies suggest that the TF method is robust for the low-dose imaging on TomoTherapy with 75% reduction of the projection views. In addition, our GPU-based solver enables the rapid image reconstruction.

REFERENCES

[1] M. J. Zelefsky, Z. Fuks, L. Happersett, H. J. Lee, C. C. Ling, C. M. Burman, M. Hunt, T. Wolfe, E. S. Venkatraman, A. Jackson, M. Skwarchuk, and S. A. Leibel, "Clinical experience with intensity modulated accelerated radiation therapy (IMRT) in prostate cancer," Radiother. Oncol. **55**, 241-249 (2000).

[2] E. B. Butler, B. S. Teh, W. H. Grant III, B. M. Uhl, R. B. Kuppersmith, J. K. Chiu, D. T. Donovan, and S. Y. Woo, "SMART (simultaneous modulated accelerated radiation therapy) boost: A new accelerated fractionation schedule for the treatment of head and neck cancer with intensity modulated radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. **45**, 21-32 (1999).

[3] T. R. Mackie, T. Holmes, S. Swerdloff, P. Reckwerdt, J. O. Deasy, J. Yang, B. Paliwal, and T. Kinsella, "Tomotherapy: a new concept for the delivery of dynamic conformal radiotherapy". Med. Phys. **20**,1709-1719 (1993).

[4] S. L. Meeks, J. F. Harmon, Jr. K. M. Langen, T. R. Willoughby, T. H. Wagner, and P. A. Kupelian, "Performance characterization of megavoltage computed tomography imaging on a helical tomotherapy unit", Med. Phys. **32**, 2673-81 (2005).

[5] K. J. Ruchala, G. H. Olivera, E. A. Schloesser, and T. R. Mackie, "Mega-voltage CT on a tomotherapy system", Phys. Med. Biol. **44**, 2597-2621 (1999).

[6] M. J. Murphy, J. Balter, S. Balter, J. A. BenComo, Jr, I. J. Das, S. B. Jiang, C. M. Ma, G. H. Olivera, R. F. Rodebaugh, and K. J. Ruchala, "The management of imaging dose during image-guided radiotherapy: Report of the AAPM task group 75," Med. Phys. **34**, 4041-4062 (2007).

[7] D. J. Brenner and E. J. Hall, "Computed tomography: An increasing source of radiation exposure," N. Engl. J. Med. **357**, 2277-2284 (2007).

[8] H. Hricak, D. J. Brenner, S. J. Adelstein, D. P. Frush, E. J. Hall, R. W. Howell, C. H. McCollough, F. A. Mettler, M. S. Pearce, and O. H. Suleiman, "Managing radiation use in medical imaging: A multifaceted challenge," Radiology **258**, 889-905 (2011).

[9] M. F. Chan, J. Yang, Y. Song, C. Burman, P. Chan, and S. Li, "Evaluation of imaging performance of major image guidance systems", Biomed Imaging Interv J. **7**, e1-7 (2011).

[10] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", IEEE Trans. Inf. Theory, **52**, 489-509 (2006).

[11] D. L. Donoho, "Compressed sensing", IEEE Trans. Inf. Theory, **52**, 1289-306 (2006).

[12] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," J. X-Ray Sci. Technol. **14**, 119-139 (2006).

[13] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," Med. Phys. **35**, 660-663 (2008).

[14] H. Y. Yu and G. Wang, "Compressed sensing based interior tomography, " Phys. Med. Biol. **54**, 2791-2805 (2009).

[15] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," Med. Phys. **37**, 5113-5125, (2010).

[16] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," Phys. Med. Biol. **56**, 3787-3806, (2010).

[17] H. Gao, J. F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," Phys. Med. Biol. **56**, 3181-3198 (2011).

[18] B. Zhao, H. Gao, H. Ding, and S. Molloi, "Tight-frame based iterative image reconstruction for spectral breast CT," Med. Phys. In Press (2013).

[19] H. Gao, R. Li, Y. Lin, and L. Xing, "4D cone beam CT via spatiotemporal tensor framelet," Med. Phys. **39**, 6943-6946 (2012).

[20] H. Gao, "Fast parallel algorithms for the x-ray transform and its adjoint," Med. Phys. **39**, 7110-7120 (2012).

[21] https://sites.google.com/site/fastxraytransform

[22] https://sites.google.com/site/spatiotemporaltensorframelet

[23] T. Goldstein and S. Osher, "The split Bregman algorithm for l_1 regularized problems," SIAM J. Imaging Sci. **2**, 323-343 (2009).

[24] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," Inverse Problems **27**, 115012 (2011).

Reduction of Artifacts Caused by High-Z Materials in Dental Spectral CBCT

Radin A. Nasirudin, Kai Mei, Petar Penchev, Ernst J. Rummeny, Martin Fiebich and Peter B. Noël

***Abstract*— The ability of photon-counting detectors to discriminate photons based on their energies has led to the development of new techniques in computed tomography (CT), in particular Spectral CT. The spectral measurements enable decomposition of the projection data into its basis components, thus providing additional information of the scanned object. This knowledge can possibly be used to improve the diagnostic quality in CT. In this work, we investigate the ability of additional spectral information to reduce metal artifacts caused by gold implants in dental cone-beam CT (CBCT). To reduce these types of artifacts we present a new algorithm, which we call Spectral driven Prior Information Reconstruction (SPIR). In step one of this algorithm a decomposition of the spectral data to determine the spatial location of the gold is performed, and in step two this information is incorporated as a prior into a penalized maximum log-likelihood reconstruction algorithm. To determine the diagnostic value of our reconstructions, subjective and objective image quality were assessed. When incorporating spectral information, a significant improvement of image quality and a significant reduction of artifacts can be reported.**

***Index Terms*—Spectral CT, Artifacts reduction, Material decomposition, Statistical image reconstruction, dental CBCT.**

I. Introduction

The advancement in detector technology has made it possible to discriminate photons based on their energy in a single scan [1]. In this particular technique, also known as Spectral CT Imaging, photon-counting detectors split the x-ray spectrum into several pre-defined energy bins, enabling the acquisition of separate CT data in each energy bin. The promising prospect of Spectral CT as the next-generation of CT systems has prompted more investigation into the various aspects of this technique [2]-[4].

This work was supported by the German Department of Education and Research (BMBF) under Grant 01EX1021D.

Radin A. Nasirudin, Kai Mei, Ernst J. Rummeny and Peter B. Noël are with the Department of Diagnostic and Interventional Radiology at Technische Universität München, Munich, Germany

Petar Penchev and Martin Fiebich are with Institute für Medizinische Physik und Strahlenschutz, Technische Hochschule Mittelhessen, Giessen, Germany.

Corresponding author: radin.nasirudin@tum.de

Our investigation focuses on Spectral CT for metal artifact reduction, especially in dental cone-beam CT (CBCT). The presence of metal objects (high Z-number materials), such as dental implants, causes the x-ray to be heavily attenuated, resulting to a reduced number of photons reaching the detector. This can led to severe streaking and dark and bright shading around the metal implant, thus degrading the diagnostic quality of the CT image. Many techniques [5]-[6] were developed to overcome this artifact, but a unique approach from Stayman et al. [7] has shown that reconstruction with prior knowledge of the material can produce superior image quality.

In this work, we investigate the ability of additional spectral information to be integrated into the reconstruction process. To reduce metal artifacts we present a new algorithm, which in step one performs a decomposition of the spectral data to determine the spatial location of the gold, and in step two incorporates that information as a prior into a penalized maximum log-likelihood reconstruction algorithm. Finally, we show and discuss the initial results from our algorithm.

II. Methods

Figure 1 illustrates an overview of our algorithm, which we call Spectral driven Prior Information Reconstruction (SPIR), where the Material Decomposition (A) and the Penalized Maximum Likelihood Iterative Reconstruction (B) are the main components.

A. Material Decomposition

As x-ray penetrates through an object, the transmitted x-ray spectrum is attenuated according to Beer's Law,

$$I = I_0 \cdot \exp^{-\int \mu(\vec{x},E)dx} \quad (1)$$

where I_0 and I are the initial and the measured intensity of the x-ray respectively. In a photon-counting detector, I can be described as the number of photons reaching the detector while I_0 the initial number of photons. The term $\int \mu(\vec{x},E)dx$ is the line integral of the attenuation coefficient along the x-ray path.

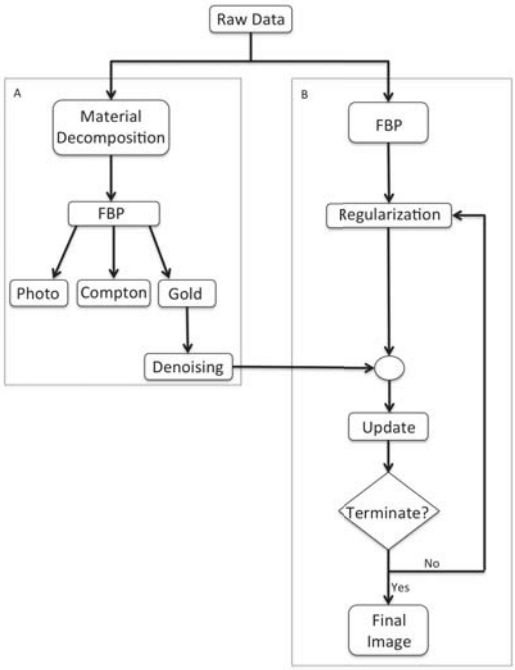


Figure 1: The workflow of our algorithm. Box A shows the material decomposition flow; box B the penalized maximum-likelihood iterative reconstruction

In the diagnostic x-ray energy range, photoelectric absorption and Compton scattering attenuate the x-rays dominantly. The photoelectric effect can be approximated by the E^{-3} energy dependence while the Compton cross section can be derived from the Klein-Nishina function [8]-[9]. In the presence of material with a high Z-number such as gold, the linear attenuation coefficient can be described [9] as

$$\mu(\vec{x},E) = A_1(\vec{x})\frac{1}{E^3} + A_2(\vec{x})f_{KN}(E) + A_3 f_{Au}(E) \qquad (2)$$

where A_1,A_2,A_3 denotes the local density of the basis function, f_{KN} the Klein-Nishina function and f_{Au} the mass attenuation coefficient of gold.

From (2), a minimum number of three x-ray intensity measurements are necessary to estimate the parameters A. For this work, we simulated a photon-counting detector with 6 energy bins (N = 6). Combining (1) and (2), the expected number of photons λ in the energy bin B_n can be described [9] as

$$\lambda_n(A_1,A_2,A_3) = \int_0^\infty S_n(E)\Phi(E)\exp^{-A_1E^{-3}-A_2f_{KN}-A_3f_{Au}} D(E)dE \qquad (3)$$
$$n = 1,....,N$$

The index n refers to n^{th} energy window, while Φ(E) denotes photon fluence and $S_n(E)$ the spectral response of the detector.

As the number of energy bins exceeds the number of attenuation basis, the system is over-determined. We used the maximum likelihood parameter estimation method to estimate the line integrals of the individual components. As it is more convenient to minimize the negative log-likelihood [9], we can express the likelihood function in terms of the measurement results ($m_1 \ldots m_N$) with respect to the parameters A as:

$$L(m_1...,m_n \mid \lambda_1(A),...,\lambda_n(A)) \cong \sum_{n=1}^{N}[\lambda_n(A) - m_n \ln \lambda_n(A)] \qquad (4)$$

This maximum likelihood technique returns in our case sinograms of photoelectric effect, Compton scattering and the attenuation of gold. We reconstructed the sinograms using a conventional FBP. Next, we performed image-processing steps on the gold image to remove noise and better determine the position of the gold implant. The gained information (density and position of the implant) is passed as a prior into the next step.

B. Penalized maximum likelihood iterative reconstruction

For the iterative reconstruction, we used a modified version of the separable paraboloidal surrogate (SPS) technique [10] with Langes regularization. This Poison-statistics-based algorithm uses the raw measurements rather than the logarithms of the data, and thus is believed to solve nonlinearity of the logarithm and handle low radiation scans.

The goal of the algorithm is to maximize a cost function Ψ, which consists of a likelihood term L

and regulation term *R*. *L* indicates how the reconstructed result matches the input sinogram; *R is* the penalty function, which reduce noise in the reconstructed slice (control of look and feel).

$$\hat{\mu} = \arg\max \Psi(\mu), \quad \Psi(\mu) = L(\mu) - \beta R(\mu) \tag{5}$$

where μ indicates the attenuation value in one pixel.

In order to maximize this function, we made use of the order-subsets version of this algorithm. Each update step is given by:

$$\mu_j^{n+1} = \left[\mu_j^n + \frac{BP[b_i \exp(-FP[\mu_j] - y_i)] - \beta \sum_k w_{jk} \dot{\psi}(x_j - x_k)}{BP[a_i y_i] + \beta \sum_k w_{jk} \ddot{\psi}(x_j - x_k)} \right]^+ \tag{6}$$

FP[] and BP[] denote the forward- and backward projection respectively; $w_{jk}\psi()$ the regulation term.

We used Langes function as a regulation term which is weighted by β. Langes function acts as a bilateral filter that eliminates noise within a threshold δ and while at the time preserving edges of the image.

$$\psi(t) = \delta^2 [|t/\delta| - \log(1 + |t/d|)] \tag{7}$$

As prior information the location and density of the gold (as determined in the previous section) is incorporated into the algorithm to enforce the correct gold attenuation. A simplified version of the algorithm can be found as pseudo code in Algorithm 1.

Algorithm 1 pseudo code

```
A ← initialize
for i = 1 to max projection data do
  for j = 1 to number of detector elements do
    A ← min Σ_{n=1}^{N} [λ_n(A) − m_n ln λ_n(A)]
    μ_metal ← BP[A]
  end for
end for
μ ← initial reconstruction
//precompute curvature:
d = BP[y · FP[1]]
for k = 0 to max iteration do
  for m = 1 to number of subset do
    l ← FP[μ + μ_metal]
    h ← b exp(−l) − y
    L ← BP[h]
    μ ← [ μ + (L − βṘ)/(d + βR̈) ]^+
  end for
end for
```

C. Monte Carlo Simulation

We simulated a phantom based on the information provided by the Phantom-Group [11] (IMP, University Erlangen-Nürnberg, Erlangen, Germany). It consisted of 32 teeth; while one of the teeth has a dental implant made out of pure gold (density 19.3 g/cm3). The photon transport mechanism was simulated using a Monte-Carlo simulator based on EGSnrc C++ class library [12-13]. The x-ray source was generated at tube voltage of 125 kV with mean spectrum energy of 55.457 keV. Six threshold levels were set at 25, 33, 51, 80, 91, and 110 keV.

III. RESULTS

Figure 2 illustrates the three basis component images; (A) photoelectric effect, (B) Compton scattering and (C) gold attenuation. One can clearly distinguish the gold implant from the other anatomic structures of the phantom such as teeth and spine.

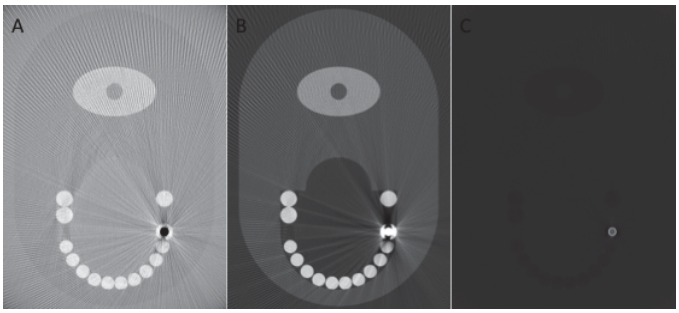


Figure 2: After material decomposition the basis sinograms are reconstructed. Image A shows photoelectric attenuation; B shows the Compton scattering and C the gold attenuation. It can be seen that, the gold implant can be distinguished from other parts of the phantom, especially the teeth.

Figure 3 presents the results of the different reconstruction. In FBP (A), it can be seen that the dental implant in the phantom causes severe streaking artifacts in the image and additional dark shadows around the dental implant as seen in the zoom-in in (D). Iterative reconstruction on the plain absorption data without prior information reduces the artifacts, as shown in (B) and (E), but reconstruction with prior information from the spectral CT delivers a significantly improved result

as seen in (C) and (F). Streaking artifacts were reduced significantly, without compromising the detail of other parts of the image such as the teeth and spine. Further, the SPIR algorithm was able to significantly reduce the dark shading around the dental implant.

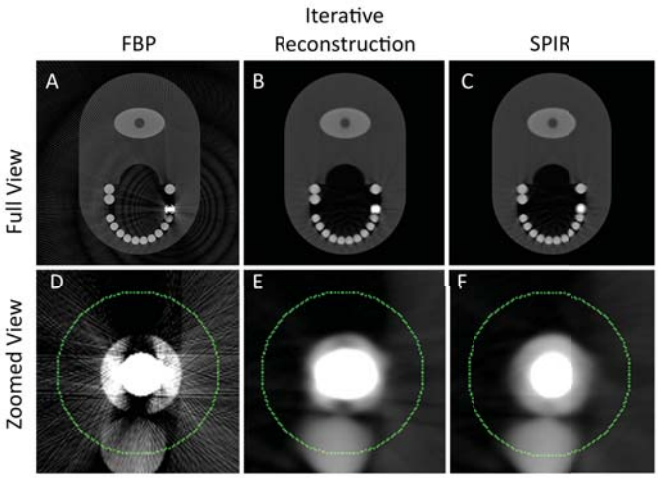


Figure 3: The row above (A-C) are the reconstructions of the whole spectral data, while the row below (D-F) is the zoom in of the area near the dental implant of each reconstructed image. The images in the middle column (B and E) are iterative reconstruction of the data without prior knowledge, while images C and F are reconstructions with SPIR. Our algorithm has shown to significantly reduce streaking artifacts and dark shadows around the implant.

In order to analyze the results quantitatively, we selected a ring-of-interest and collected pixel values along the circumference, as indicated in green in Figure 3. In Figure 4, one can see the pixel values (which are collected along the green ring in figure 3) of different components (tissue and teeth) around the implant. The SPIR algorithm reflects the actual theoretical value better in comparison to the other algorithms.

IV. Conclusions

We have demonstrated that the combination of spectral information and statistical reconstruction can significantly improve image quality, especially in the presence of high Z-materials. . While the SPIR algorithm has proven to reduce streaking artifacts in dental cone-beam CT, our algorithm could be extended to other parts of body with metal prosthesis or implants such as the lower extremity or spine. While metal artifacts are one of the most common artifacts in CT, we believe that the information provided by Spectral CT can be useful in overcoming not only this image quality issues. One could foresee the integration of further information to improve the diagnostic image quality while possibly reducing the radiation dose to the general patient population.

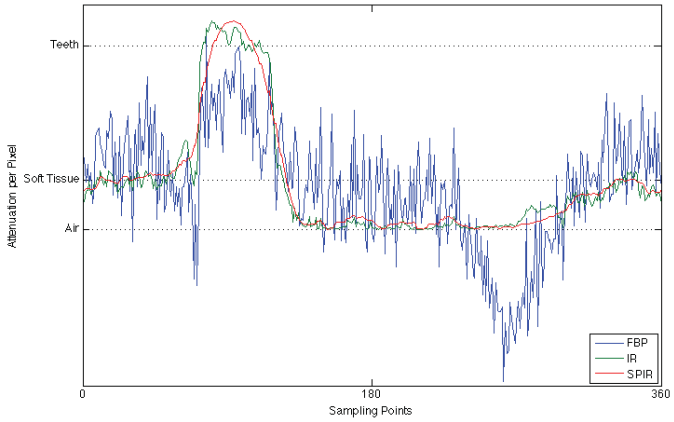


Figure 4: The graph above shows the pixel values around the circle drawn on the reconstructed images. (see Figure 3). The theoretical attenuation values for air, soft tissue and teeth is treated as the ground truth. It can be seen that our algorithm reflects the true value better in comparison to other reconstruction.

References

[1] Fischer, P., & Helmich, A. (2000). A photon counting pixel chip with energy windowing. *IEEE Transactions on Nuclear Science*, *47*(3).

[2] Roessl, E., & Proksa, R. (2007). K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors. *Physics in medicine and biology*, *52*(15), 4679–96.

[3] Wiegert, J., Engel, K. J., & Herrmann, C. (2009). Impact of scattered radiation on spectral CT. *Proceedings of SPIE*, *7258*, 72583X–72583X–10.

[4] Schmidt, T. G. (2009). Optimal “image-based” weighting for energy-resolved CT. *Medical Physics*, *36*(7), 3018. Retrieved from

[5] Kalender, W. A., Hebel, R., Ebersberger, J. (1987, August). Reduction of CT artifacts caused by metallic implants. *Radiology*, vol. 164, pp. 576-7.

[6] Wang, G., Snyder, D. L., O’Sullivan, J. a, & Vannier, M. W. (1996). Iterative deblurring for CT metal artifact reduction. *IEEE transactions on medical imaging*, *15*(5), 657–64.

[7] Stayman, J. W., Otake, Y., Prince, J. L., Khanna, a J., & Siewerdsen, J. H. (2012). Model-based tomographic reconstruction of objects containing known components. *IEEE transactions on medical imaging*, *31*(10), 1837–48.

[8] Alvarez, R. E., & Macovski, A. (1976). Energy-selective reconstructions in X-ray computerized tomography. *Physics in Medicine and Biology*, *21*(5), 733–744.

[9] Schlomka, J. P., Roessl, E., Dorscheid, R., Dill, S., Martens, G., Istel, T., Bäumer, C., et al. (2008). Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography. *Physics in medicine and biology*, *53*(15), 4031–47.

[10] J.A. Fessler. Statistical image reconstruction methods for transmission tomography. *Handbook of Medical Imaging*, 2:1–70, 2000.

[11] Watzke, O., “Jaw Phantom,” Insititute of Medical Physics, Friedrich-Alexander- University Erlangen-Nü rnberg, Erlangen, Germany, http://www.imp.uni- erlangen.de/phantoms/jaw/jaw.htm

[12] Kawrakow, I., Mainegra-Hing, E., Tessier, F., Walters, B. R. B., “The EGSnrc C++ class library,” NRC Report PIRS-898 (rev A), Ottawa, Canada, 2009

[13] Noel, P. B., Penchev, J. M., Renger, B., Rummeny, E. & Fiebich, M. “A Monte-Carlo Based Software-Bench for Performance Characterization of CT Reconstruction Algorithms,” In Joint AAPM/COMP Meeting. Vancouver, 2011

Grid-free backprojection-maximization algorithm for 3D imaging using a vehicle-mounted coded aperture gamma camera

Jonathan S. Maltz, Lucian Mihailescu, Donald L. Gunter, Tim Aucott, Grant T. Gullberg, and Kai Vetter

***Abstract*— Vehicle-based radioactive source detection systems play an important role in contemporary strategies for detecting illicit nuclear materials in civilian environments. Three-dimensional imaging systems, such as coded aperture (CA) cameras, theoretically offer improved sensitivity and localization performance over methods based on 2D far-field imaging, or solely on source strength. However, the imaging grid needed to represent all possible source locations to an acceptable resolution makes fully 3D imaging difficult in practice. We present a grid-free method of locating sources based on maximizing the backprojection function using stochastic optimization. When multiple sources are present, the algorithm sequentially locates the strongest sources by penalizing the objective function near the positions of previously detected sources. Test data are obtained using a truck-mounted CA camera with a 10×10 array of 10×10×5 cm^3 NaI detectors to image a 1 mCi ^{137}Cs source and a 2 mCi ^{131}I source, at respective distances of closest approach of 10 m and 25 m. To enable receiver-operating characteristic analysis of the real data, we obtained 53 minutes of background data while driving through an urban environment. In a simulation study based on the real data, the algorithm is able to locate the 1 mCi ^{137}Cs source to within 27 cm of its true position. When the algorithm is applied to real data, perfect detection with no false alarms is achieved for both sources. Execution speed 8 ms/event/core on a contemporary CPU, making the algorithm practical for real-time deployment at realistic count rates and vehicle velocities. The output of the algorithm can be used to design a variable resolution grid for subsequent quantitative image reconstruction.**

***Index Terms*— coded aperture imaging, image reconstruction, nuclear source detection**

I. Introduction

Vehicle-mounted gamma imaging systems are an important part of a strategy to detect the presence of illicit nuclear material. Most existing implementations detecting sources by analyzing 2D far-field images obtained using coded aperture and Compton camera systems. These images are then used to estimate the 3D locations of point sources.

Without fully 3D imaging, it is not possible to make optimal use of the parallax gained along the trajectory of a moving vehicle. In far-field imaging, all sources along each ray between the detector and the source sphere contribute to each of the 2D bins, leading to generally inferior signal-to-noise ratio when compared to the case where the image support is partitioned into 3D voxels. System detection performance should thus be improved in fully 3D implementations. However, quantitative 3D reconstruction is complicated by the vast grid required for the image, and incomplete angular tomographic sampling due to the irregular truck trajectory. Indeed, full quantitation may often be impossible.

The algorithm presented here is a pragmatic semi-quantitative approach towards solving the 3D source imaging problem in real time. The output of the algorithm is a list of source locations based on the maxima of the 3D backprojection function. This information may then be used to construct either a mesh (e.g., octree), or point cloud, so that the full reconstruction problem can be solved in an efficient way.

We formulate the method and then demonstrate that it can accurately locate sources in 3D under realistic urban background conditions. We also show that on-line operation of the algorithm is feasible using contemporary computer hardware.

This work has been supported by the US Department of Homeland Security, Domestic Nuclear Detection Office, under competitively awarded contract/IAA HSHQDC-08-X-00832 and by the U.S. Department of Energy, Office of Science, under Contract DE-AC02-05CH11231. This support does not constitute an express or implied endorsement on the part of the Government.

All authors are with the Nuclear Science Division of Lawrence Berkeley National Laboratory. TA, and KV are also with the Department of Nuclear Engineering, University of California at Berkeley.

II. Methods

We begin by formulating the backprojection (BP) function, upon which the optimization objective function is based.

A. The backprojection function

The algorithm makes use of two co-ordinate systems. The world co-ordinates are given by $\mathbf{r}_w = (x, y, z)$. A position in the camera frame is $\mathbf{r}_c = (u, v, w)$.

Our approach involves finding the maximum of the BP function of events obtained while the camera is moving. To perform this optimization, candidate source points are generated stochastically in the world frame. For each potential source point, the BP function is evaluated only at that position. Since the measurements are recorded in the camera frame, we need to a method of converting world co-ordinates to camera co-ordinates.

The camera-to-world transformation matrix is:

$$\mathbf{T} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & u \\ R_{21} & R_{22} & R_{23} & v \\ R_{31} & R_{32} & R_{33} & w \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where the R_{**} are derived from the vehicle roll, pitch and yaw (bearing, road forward inclination, and road lateral inclination in the case of a truck). Representing the $\mathbf{r}_*$ in homogeneous co-ordinates as $\mathbf{r}'_* = [\mathbf{r}^T\ 1]^T$, we have $\mathbf{r}'_c = \mathbf{T}^{-1}\mathbf{r}'_w$, which is the desired world-to-camera transformation.

The measured data provide $\mathbf{q}_c^n$, the location of the nth event recorded in the detector. Let $\mathbf{p}_c$ represent a candidate point at which the BP value is to be calculated. For event n, the position of this point in the camera frame is $\mathbf{p}_c^{n\prime} = \mathbf{T}_n^{-1}\mathbf{p}'_w$.

The matrix $\mathbf{T}_n$ is the camera-to-world transformation that applies at the time event n is measured. The event-normalized BP value is given by:

$$b(\mathbf{p}_c) = w(E)\,\frac{1}{(N_2 - N_1)}\sum_{n=N_1}^{N_2}\int_{\mathbf{r}_c\in\mathbf{\Gamma}^n} d\,\mathbf{r}_c\,\mathrm{e}^{\,-\boldsymbol{\mu}(\mathbf{r}_c,E)\,\mathbf{r}_c}, \tag{2}$$

where $\mathbf{\Gamma}^n = \{\mathbf{r}_c \mid \mathbf{r}_c = \mathbf{q}_c^n + \lambda(\mathbf{p}_c^n - \mathbf{q}_c^n),\ \lambda \in [0,\,1]\}$, and $\boldsymbol{\mu}(\mathbf{r}_c, E)$ is the volume containing the attenuation coefficients for the mask and truck body at energy E. N_1 and N_2 are the bounds of the event window, which are based on the event partition strategy used (see Section II-C below).

In out implementation, we use the energy weighting factor $w(E)$ to effect background subtraction. Events in the photopeak window of length ΔE_p are given a weight of unity. For events in window k of K background windows, having width ΔE_b^k, the weight is calculated as:

$$w^k = -\frac{1}{K}\frac{\Delta E_p}{\Delta E_b^k}. \tag{3}$$

This gives equal weight to each of the background windows chosen.

B. Maximizing the backprojection function

The objective of the optimization is to solve:

$$\hat{\mathbf{p}}_{\mathbf{w}} = \arg\max_{\mathbf{p}'_w} (t_2 - t_1)\, b(\mathbf{H}\mathbf{T}_n^{-1}\mathbf{p}'_w), \quad n = N_1, \ldots, N_2, \quad (4)$$

where $\mathbf{H} = [\mathbf{I}_{3\times3}\ \mathbf{0}_{3\times1}]$ simply converts from homogeneous to natural co-ordinates. t_1 and t_2 are the times at which events N_1 and N_2 are recorded, respectively. The objective function can be viewed as assessing the amount of coherence in the backprojection, normalized by the mean count rate in the event window.

Since the optimization problem (4) typically has multiple local maxima, we use simulated annealing to find a solution [1].

C. Selection of events

In standard tomography, we would solve (4) using all available measured events. In the case of a truck, the path of the camera is generally irregular and so only a subset of recorded events are relevant to the current field-of-view (FOV). A simple way of addressing this issue is to assume the trajectory is piece-wise linear. The FOV can then be calculated based on the initial and final positions of the vehicle for each linear part of the trajectory.

The FOV of the stationary camera is defined to include only those events that reach the detector after traveling through the mask structure. The centers of the detector and the mask have (u, v, w) co-ordinates $(0, 0, 0)$ and $(0, 0, l_s)$, respectively. The u-axis is oriented along the forward direction of truck travel, and the v-axis increases away from the ground. The tangent of the bounding view angle is given by $\tan(\theta_w) = l_s/(w_m + w_d)$, where w_m and w_d are the half-widths of the mask. Let $\mathbf{s}_1$ represent the start of a linear segment, $\mathbf{h}$ be a unit vector containing the current heading, and $l_{\max}$ denote the maximum expected distance to a source. The limit of the event trailing window includes events that occur only after the truck passes the point: $\mathbf{s}_t = \mathbf{s}_1 - l_{\max}\cot(\theta_w)\,\mathbf{h}$. For retrospective analysis, the forward limit is similarly $\mathbf{s}_f = \mathbf{s}_2 + l_{\max}\cot(\theta_w)\,\mathbf{h}$, where $\mathbf{s}_2$ is the end point of the current linear segment. For prospective on-line use, all data acquired after driving past $\mathbf{s}_t$ are used.

The linear constraints on the solution associated with the FOV bounds prevent the simulated annealing algorithm from accepting parameter states corresponding to out-of-bound source positions.

In many source detection settings, 3D optical camera information and geographic information systems (GISs) may be used to adaptively set $l_{\max}$ in a rational way. This can improve performance by limiting the event trailing window to the most relevant events.

D. Integration with a GIS

The truck GPS co-ordinates, as well as the solution locations and objective function maxima for each time window, are fed to the Google Earth (Google Inc., Mountain View, CA) GIS using a keyhole markup language (kml) file that encodes the vehicle trajectory as a "tour", which can be viewed live or reviewed at any time.

III. Evaluation

A. Simulated data

The algorithm is applied to Monte Carlo-generated data modeled realistically on an actual acquisition involving driving by a 1 mCi ^{137}Cs source with a distance-of-closest-approach of 10 m. The data are acquired by the Mobile Imaging and Spectroscopic Threat Identification (MISTI) platform described in [2]. An schematic illustration of this truck appears in Figure III-A.

The source is moved in the camera reference frame in order to simplify the simulation process. Background sources are modeled as rays randomly intersecting the detector surface originating from the random vertices of tesselated half-sphere centered at and in front of the detector. The half-sphere has 512 vertices and a radius of 5 m. Figure 2 shows the cumulative counts attributable to the source and background, as well as the aggregate curve that matches the measured cumulative-counts-vs-time curve.

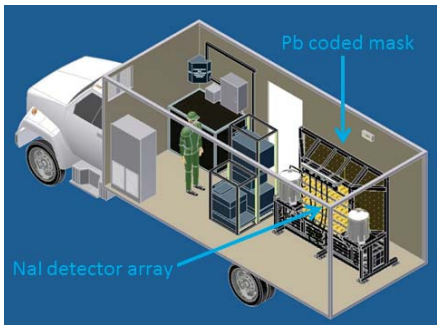


Fig. 1. The MISTI truck contains a coded aperture camera with a 10×10 array of 10×10×5 cm^3 NaI detectors. The detector plane is oriented parallel to the right side of the vehicle. Image reproduced from [2] and adapted.

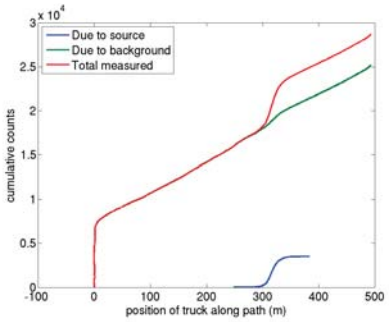


Fig. 2. These curves of cumulative counts versus vehicle position are derived from the measurement data using counts from the photopeak and background energy windows. They are used to produce a realistic Monte Carlo simulation of the image acquisition process.

A custom Monte Carlo code, named DOSxyznrc_compt, is used to perform the simulation. It is a version of the EGSnrc user code DOSxyznrc [3] that we have modified to accommodate 3D phase space input and particle tracking for Compton cameras. For use with coded aperture imagers, the positions of first energy deposition in the detectors are used as the event locations within the detector.

To define the event windows for optimization, we set $l_{\max}$=50 m. We run the algorithm in prospective mode to simulate its on-line application. The optimization problem (4) is solved for 11 adjacent event windows over the 500 m truck trajectory. Overlap between successive windows is set at $l_{\max}/5$.

Since the simulated source and background events are all within the photopeak window, we set $w(E) = 1$ in (2).

B. Real data

We examine two test scenarios, A and B. In both cases, the MISTI truck is driven along a straight road at 15 mph. In Scenario A, the truck drives past a 1 mCi ^{137}Cs source at a distance-of-closest-approach (DoCA) of 10 m. In Scenario B, we drive past a 2 mCi ^{131}I source at a DoCA of 25 m.

We recorded 10 and 12 source runs for Scenarios A and B, respectively.

To enable us to determine the receiver-operating-characteristic (ROC) for detection, we obtained background data driving around an urban area for 53 minutes.

Background energy window subtraction is effected using the windows listed in Table I.

TABLE I
PHOTOPEAK AND BACKGROUND WINDOWS (IN keV) USED TO IMPLEMENT BACKGROUND SUBTRACTION DURING LIST-MODE BACKPROJECTION.

Source isotope	Photopeak window	Background window 1	Background window 2
^{137}Cs	635.5–691.8	540.0–580.0	740.0–800.0
^{131}I	349.4–380.4	296.9–323.2	406.7–439.9

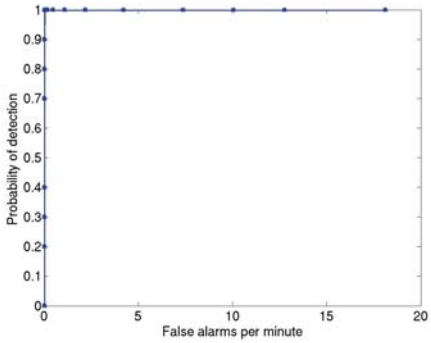


Fig. 3. ROC curve for detecting the 1 mCi ^{137}Cs source at a minimum range of 10 m. The marker "x" indicates a threshold value sample used to form the curve. The area under this curve is unity, indicating perfect detection at zero false alarm rate.

As in the simulation study, we set $l_{\max}$=50 m, with 10 m of overlap between adjacent windows.

IV. RESULTS

A. Simulated data

Figure 5 contains snapshots from a video recreation of the view from the truck. The GPS co-ordinates recorded during the measurement interval are fed to the GIS along with the markers indicating the solutions to (4) obtained for each of the 11 event windows. Figure 6 shows the maximum values of b for each of the event windows. The true source is located in window 8, where $b = 0.753$, which is 41% greater than the next highest value. It is thus easy to set a threshold that will allow detection of the source over background.

Figure 7 shows the BP function for event window 8 in the plane containing the true source that is oriented parallel to the detector face. The maximum of the objective function is very close to the true solution (it is in fact 27 cm away).

The mean execution speed of the algorithm is 8 ms/event/core on an Intel Xeon 5520 processor operating at 2.27 GHz. (Simulated annealing is a highly parallel algorithm, so execution time scales as the inverse of the number of cores.) The C-code is compiled using gcc version 4.6.0 with the `-O3` optimization setting.

B. Real data

The ROC curves for the ^{137}Cs and ^{131}I sources appear in Figures 3 and 4, respectively. ROC analysis shows that perfect detection performance is possible at a zero false alarm rate.

V. CONCLUSION

The proposed algorithm performs very well when applied to imaging the relatively strong sources examined here. We intend to extend our analysis to data generated by considerably weaker sources, such as 350 mCi ^{131}I sources at 25 m minimum range.

In the simulation run, when the ground truth source position was known exactly, the 1 mCi ^{137}Cs at 10 m was located to within 27 cm of its true position.

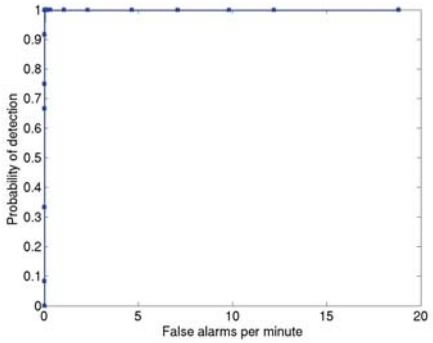


Fig. 4. ROC curve for detecting the 2 mCi ^{131}I source at a minimum range of 25 m. The marker "x" indicates a threshold value sample used to form the curve. The area under this curve is unity, indicating perfect detection at zero false alarm rate.

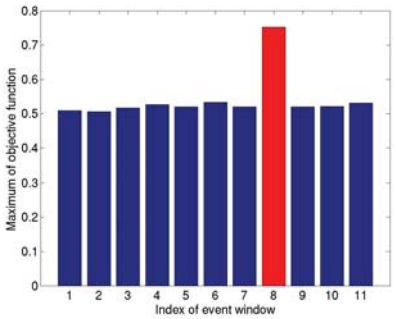


Fig. 6. Bar plot of maximized objective function values for each of the 11 event windows. The solution found for window 8 is the true solution, and corresponds to an objective function value that is 41% higher than the next highest value.

When applied to real data acquired using the MISTI truck, perfect detection at zero false alarm rate was demonstrated for both sources.

Computational requirements are modest in the context of contemporary parallel computer hardware capabilities, making it suitable for on-line use.

This algorithm should be considered only a first step in fully 3D source imaging. It is useful for initial detection of sources, and as a basis for constructing multiresolution reconstruction grids. Backprojection alone has limited resolving capabilities, and proper solution of the inverse problem will produce more useful images, particularly in cases where the vehicle trajectory produces higher quality tomographic datasets. In practice, this may require driving around the putative source location (as identified using this algorithm, or otherwise) in order to obtain more complete angular sampling.

REFERENCES

[1] A. L. Ingber, "Very fast simulated re-annealing," *Journal of Mathematical Computer Modelling*, vol. 12, no. 8, pp. 967–973, 1989.
[2] E. A. Wulf, B. F. Phlips, W. Neil Johnson, B. Leas, and L. J. Mitchell, "MISTI imaging and source localization," in *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pp. 2413–2417, oct. 2008.
[3] D. Rogers, I. Kawrakow, and B. Walters, "DOSXYZnrc users manual," Tech. Rep. NRCC Report PIRS-794 Rev B., Ionizing Radiation Standards National Research Council of Canada, 2009.
[4] B. Horn, R. Lanza, J. Bell, and G. Kohse, "Dynamic reconstruction," *Nuclear Science, IEEE Transactions on*, vol. 57, pp. 193 –205, Feb. 2010.

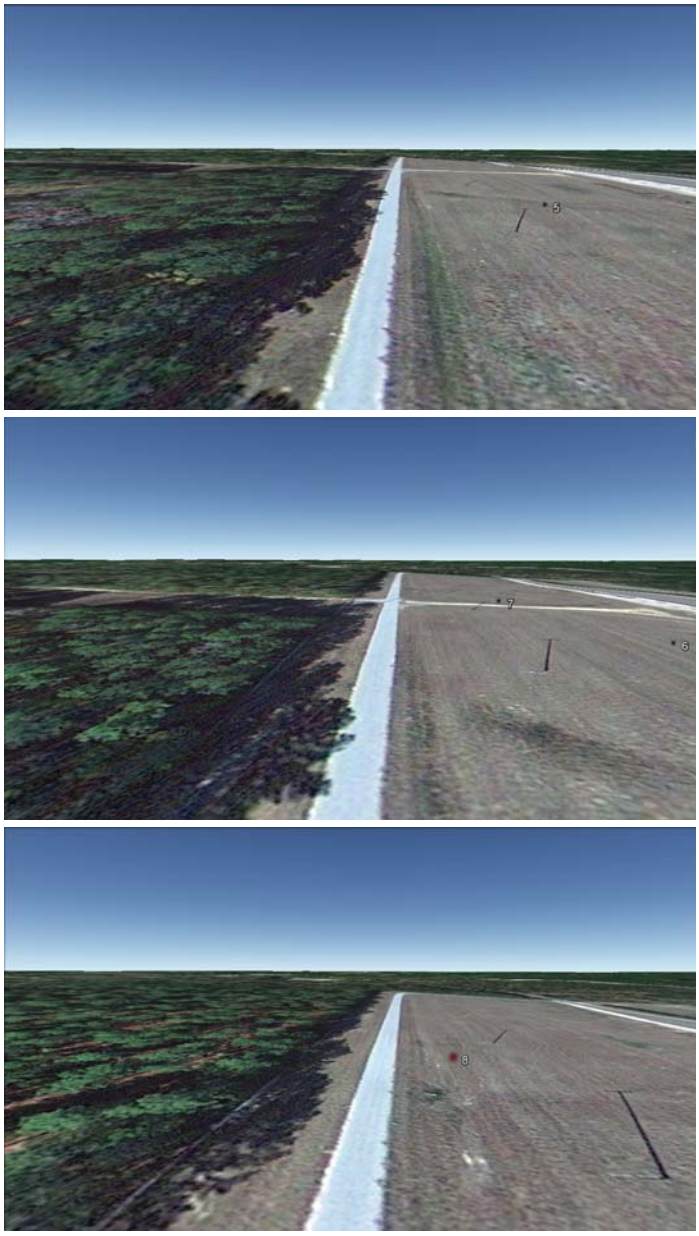


Fig. 5. From top to bottom, the figures illustrate snapshots of a virtual rerun of the actual path taken by the truck. In the top and middle figures, solutions associated with low values of the objective function (true negatives) are shown as black dots. In the lower figure, the actual source position (true positive) is indicated by the red dot. The number labels indicate the index of the event time window used in the calculation

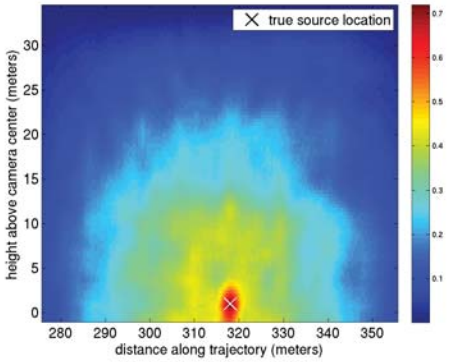


Fig. 7. Backprojection function on plane through true source location, formed from events within window 8. A strong local peak is visible in the immediate vicinity of the true source.

Volume Splitting Based Multi-GPUs Implementation for 3D List-Mode PET Reconstruction

Zakaria Bahi, Julien Bert and Dimitris Visvikis

***Abstract*—In PET imaging, one main obstacle in obtaining a fully quantitative list-mode reconstruction in a run time compatible for clinical environment is the computation burden. Among parallelization and optimization architectures, graphics processing units (GPUs) represent today a powerful accelerators, especially for medical image processing. However, a reconstruction on a single GPU is insufficient to handle all the corrections needed (patient and scanner) with a compatible time for a clinical use. Multi-GPUs are now becoming the best solution to go further for higher performance computing. Only one method has been proposed recently on Multi-GPUs context for list-mode PET image reconstruction. This method is not optimized for a high-resolution reconstruction context. In this work we propose a new Multi-GPU acceleration method that optimizes the communication cost between GPUs. The results have shown that we obtained a linear performance computing scalability, that increases on higher resolutions, with a speedup factor of x4.4 between 1 and 4 GPUs.**

***Index Terms*—PET reconstruction, Multi-GPUs, high performance computing, volume splitting.**

I. Introduction

IN Positron Emission Tomography (PET) reconstruction, List-mode (LM) technique has several advantages [1] over conventional approaches. Advantages include preservation of the high frequency and high spatial resolution of the acquired data, which in turn facilitates a better handling of dynamic processes, as well as a finer image spatial resolution. Such reconstruction, suffers of high computational time cost, especially for a fully quantitative reconstruction within clinically relevant execution times. Recently, graphics processing units (GPUs) are becoming the most suitable solution to resolve computational time cost problems, thanks to its highly parallel architecture. Within this context a few PET image reconstruction implementations have been recently proposed on GPU [2]–[4]. However a PET list mode based image reconstruction that handles all necessary corrections and be executed on single GPUs, will still not be fast enough to reach a run time compatible with clinical environment requirements.

A solution consists of going further by using multiple GPUs that allows a higher performance computing. Until now, just one work was proposed on multi-GPUs for PET list-mode reconstruction using the LM-OSEM algorithm [5]. This method is based on subset splitting into chunks, each one on a GPU, meaning that every GPU processes just a small number of lines-of-response (LORs) leading to low performance computing. This method also requires maintaining the whole reconstruction volume in each graphical card memory, which is a limitation in terms of handling large images and thus allowing a high resolution reconstruction.

Z. Bahi, J. Bert and D. Visvikis are with the UMR 1101 INSERM, LaTIM, CHRU Brest, France (e-mail: zakaria.bahi@telecom-bretagne.eu).

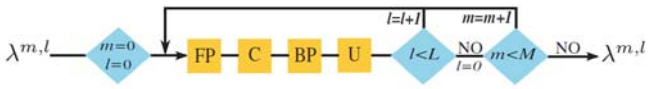


Fig. 1. List-mode OSEM algorithm diagram, where *FP* is forward projection operation, *C* is the correction operation, *BP* is back projection operation, *U* is the update operation, *l* is the subset number, and *m* is the iteration number.

Instead of splitting subsets, in this work we propose a new method that splits the volume directly and loading each portion on a GPU. This technique facilitates the highest occupancy of the GPU by processing the full subset of LORs on each one. In addition, as the reconstructed volume is scattered in pieces on each GPU memory it is possible to handle larger images than the previously proposed subset splitting method.

II. Materials and Methods

A. LM-OSEM reconstruction

List-Mode Ordered Subsets Expectation Maximization (LM-OSEM) PET reconstruction algorithm [6] was introduced to accelerate the reconstruction time of Maximum Likelihood Expectation Maximization algorithm (MLEM) without degradation of image quality. This slight modification of MLEM uses subsets of the entire data set for each image update in the form:

$$\lambda_j^{m,l+1} = \frac{\lambda_j^{m,l}}{\sum_{i=1}^{I} w_{ii} P_{ij}} \sum_{k \in S_l} P_{i_k j} \frac{1}{\sum_{b=1}^{J} P_{i_k b} \lambda_j^{m,l}} \qquad (1)$$

Where $\lambda_j^{m,l}$ is the image estimation in voxel *j (j=1,…, J)* at the m^{th} iteration and the l^{th} list-mode subset *l (l=1,…, L)* by dividing the data space into *L* subsets, and P_{ij} is the probability of an emission from voxel *j* being detected along LOR *i* (often referred as the 'system matrix'), and w_{ii} is the sensitivity matrix. i_k refers to the LOR along which the k^{th} list-mode event is detected in the subsets number *L*. Due to the system matrix high complexity and enormous size, this matrix may be decomposed in multiple sub-matrices, each accounting for different physical effects of the detection process. Within the LM context, the sub-matrix modeling the probability density function (PDF) for each line-of-response (LOR) is computed on-the-fly using a projector.

As it is shown in Fig. 1, the LM-OSEM reconstruction algorithm consists of four main sequential stages: forward projection, correction, back projection and the update stage. The forward projection computes the sums of voxels values

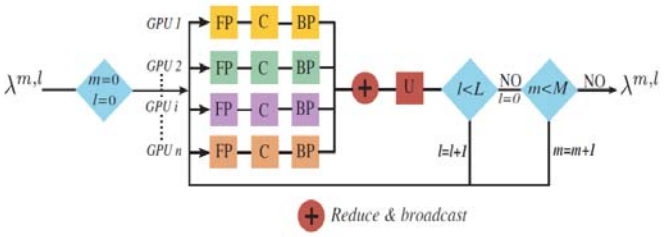


Fig. 2. Subsets splitting method flow chart.

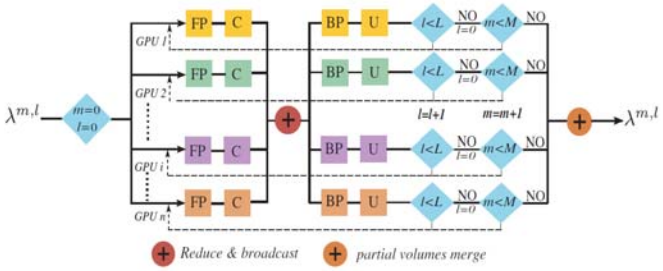


Fig. 3. Volume splitting method flow chart.

on each LOR in a single value Q, while the correction stage computes the correction values C by reversing the Q values i.e. $C = 1/Q$. This correction step will serve us in our method structure and decomposition. The back projection step traces the LORs using the correction values C of each LOR. Therefore, the image is updated after each processed subset defining a complete iteration, and repeat it until convergence.

B. Subsets parallelization

The subsets splitting method [5] is based on subset parallelization. By splitting every subset into N chunks, where N is the GPUs number, each GPU process a chunk in a parallel way. As it is known, LM-OSEM reconstruction algorithm comprises four stages: forward projection, correction computing, back projection and finally the update stage. In this subsets splitting method, as it is shown in Fig. 2, the processing order does not change for any subset chunk, so for each subset the processing algorithm is performed in the simple order independently and synchronized over all the GPUs, until the update computing step, where this stage could not be performed unless with all LORs back projection results of the subset. Thus a merge operation is required over all GPUs nodes of each subset, which we call *reduce*. This merging consists of adding together results from the different GPUs into one specific GPU, called the main node. This GPU will be in charge of performing the update stage and broadcasting its result to the other GPUs in order to start a new iteration. This communication time cost has a negative impact significantly on the global reconstruction time, caused by the reduce process on the whole image volume size over each subset.

C. Volume parallelization

In multi-GPUs PET reconstruction, there are two main obstacles. The first issue is the cost of the communications between GPUs occurring for each subset, and the second one is associated to the bottleneck of the GPU architecture, which is the high latency in accessing the global memory. We propose resolving both issues by a method that parallelizes the reconstructed volume instead of the subsets. This method is based on splitting the image volume into N portions according to the number of GPUs, i.e. each portion for one GPU. The GPU reconstruction run time is led by the memory access to read values from the volume. By splitting the volume into different portions, each GPU will decrease its memory bottleneck and run faster. As it is shown in Fig. 3, on each subset computation, all GPUs are synchronized while they are processing in an independent way the same subset. In this method, the corrections were computed partially from a volume chunk based on the same subset LORs. Therefore, a reduce summation operation on all the LORs correction values is required for all GPUs on each subset, followed by broadcasting the resulting correction values to all GPUs. Because of the small size of the correction vector (number of LORs of a subset), the communication cost is very low comparing to the subsets splitting method, which has to be performed on the whole volume. After computing the back projection step independently on each partial volume, the update computation on this volume is done as well for each GPU. The updating stage should be faster than a subset splitting method because each partial volume is computed in a parallel way. Finally, at the end of all iterations, a last reduce is required to merge and recompose all GPUs partial volumes to a final and a complete one.

D. Implementation

The algorithm implementation is based on Nvidia GPU parallel computing platform CUDA. This programming model enables powerful increases in computing performance by harnessing the power of GPU. The Nvidia GPUs is composed of thousand of threads, each one representing a data unit, in other words a piece of data that one wish to process. All threads will execute the same program code, called kernel, in parallel on the different streaming processing, representing individual processor units. We choose a strategy that consists of executing a thread per LOR (i.e. LOR chunk of a GPU data) in an independent way. For multi-GPUs, we need a specific communication interface between GPUs, Message Passing Interface (MPI) is a specification of message passing libraries, used for distributed memory environment in parallel programming models. All communications of reduce and broadcast operations are handled with MPI, which is the suitable communication interface for such distributed systems.

E. Evaluation study

1) Datasets: List-mode datasets were generated by a Monte Carlo simulated PET scanner (Philips GEMINI) [7] using

TABLE I
SUBSETS SPLITTING RECONSTRUCTION RUN TIME [S] FOR 2 GPUS

Resolution	$4^3\ mm^3$	$2^3\ mm^3$	$1^3\ mm^3$
Forward projection	5.2	24.24	201.7
Back projection	13.44	46.24	389
Comm. update	0.024	0.22	2.16
Update	$3.5{\cdot}10^{-5}$	$5{\cdot}10^{-5}$	$5.4{\cdot}10^{-5}$
Total comm.	0.072	0.67	6.48
Total run-time	57.43	212.1	1778.4
Comm/run-time	0.12%	0.31%	0.36%
Speedup / 1GPU	$\times 1.99$	$\times 1.88$	$\times 1.87$

TABLE II
VOLUME SPLITTING RECONSTRUCTION RUN TIME [S] FOR 2 GPUS

Resolution	$4^3\ mm^3$	$2^3\ mm^3$	$1^3\ mm^3$
Forward projection	5.44	21.04	200.08
Comm. correction	0.072	0.072	0.072
Back projection	14.24	44	370.4
Update	$17.6{\cdot}10^{-5}$	$17{\cdot}10^{-5}$	$21.6{\cdot}10^{-5}$
Comm. volume	0.001	0.01	0.08
Total comm.	0.2	0.2	0.2
Total run time	59	195.32	1713.8
Comm / run-time	0.33%	0.1%	0.01%
Speedup / 1GPU	$\times 1.94$	$\times 2.05$	$\times 1.94$

TABLE III
SUBSET SPLITTING RECONSTRUCTION RUN TIME [S] FOR 4 GPUS

Resolution	$4^3\ mm^3$	$2^3\ mm^3$	$1^3\ mm^3$
Forward projection	2.8	11.36	97.6
Back projection	6.72	23.12	194.64
Comm. update	0.072	0.58	5.04
Update	$34.4{\cdot}10^{-5}$	$45{\cdot}10^{-5}$	$35{\cdot}10^{-5}$
Total comm.	0.22	1.77	15.12
Total run-time	28.7	105.2	891.84
Comm/run-time	0.8%	1.68%	1.7%
Speedup / 1GPU	$\times 3.88$	$\times 3.8$	$\times 3.73$

TABLE IV
VOLUME SPLITTING RECONSTRUCTION RUN TIME [S] FOR 4 GPUS

Resolution	$4^3\ mm^3$	$2^3\ mm^3$	$1^3\ mm^3$
Forward projection	3.44	13.44	100.56
Comm. correction	0.16	0.16	0.16
Back projection	6.56	18.48	153.36
Update	$1.8{\cdot}10^{-5}$	$2.3{\cdot}10^{-5}$	$2.3{\cdot}10^{-5}$
Comm. volume	0.003	0.03	0.21
Total comm.	0.5	0.5	0.5
Total run time	30.48	96.24	762.4
Comm / run-time	1.64%	0.52%	0.06%
Speedup / 1GPU	$\times 3.75$	$\times 4.16$	$\times 4.36$

GATE [8]. A list-mode dataset of about $17{\cdot}10^6$ true unscattered coincidences was simulated for NEMA-IEC phantom, composed of a warm water cylinder containing four hot spheres with diameters 10, 13, 17 and 22 mm and two cold spheres with diameters 28 and 37 mm. From this phantom, the corresponding attenuation map was built based on attenuation coefficients provided by a tissue atlas at 511keV.

2) Reconstruction: The reconstruction went with 3 different resolutions with voxel sizes of $4\times4\times4\ mm^3$, $2\times2\times2\ mm^3$ and $1\times1\times1\ mm^3$ for reconstruction volume sizes of $141\times141\times45$, $283\times283\times89$ and $565\times565\times177$ voxels respectively. We used a Gaussian projector [9], [10] with a stationary point spread function set with a full width at half maximum at $2.4\times3.6\ mm^2$. The 3D LM-OSEM algorithm was used on a convergence point of 3 iterations and 8 equal subsets with about $2{\cdot}10^6$ LORs for each subset. The reconstructions were repeated by using 1, 2 and 4 GPUs for both splitting methods. We used two bi-GPU graphical cards (total of 4 GPUs) of NVIDIA GTX 590 with 512 CUDA cores and 1.5 GB of global memory on each GPU.

III. RESULTS

In this section we present the results of both subsets splitting and the volume splitting methods reconstructions. Tables I, II and III, IV show the reconstruction performances for both methods on 2 and 4 GPUs respectively. By looking at the reconstruction performances on 2 GPUs, we observe two main difference points for the two methods, communication cost and run time. For subsets splitting method in table I, the communication cost is increased according to the reconstruction image volume, that started at *0.072s* for 4^3mm^3 resolution, *0.67s* for 2^3mm^3 resolution and finally *6.48s* for 1^3mm^3 resolution, with an increasing factor of more than $\times9$. This communication cost increase impacts the run time speedup with $\times1.87$ factor for 1^3mm^3 resolution, instead of a theoretical factor of $\times2$. But in the other side, in table II for the volume splitting method, we see that the communication cost is constant on *0.2s* for all the 3 different resolutions, which leads to a linear speedup of a $\times1.94$ for 1^3mm^3 resolution.

For 4 GPUs reconstruction performances, we focus always on the same two points of communication cost and speedup. We still get the same observations for subsets splitting method on the communication cost that still increasing by resolution improvement, of *0.22s*, *1.77s* and *15.12s* for $4^3\ mm^3$, $2^3\ mm^3$ and $1^3\ mm^3$ respectively. The high communication cost has a worst impact on the speedup more than the reconstruction on 2 GPUs, with a final speedup loss of $\times3.73$ for $1^3\ mm^3$ resolution. Volume splitting method still has the same advantage of keeping a constant communication cost of *0.5s* for all the 3 different resolutions, which allows reaching a better speedup than the architecture parallelization of $\times4$ with factors of $\times4.16$ for $2^3\ mm^3$ and $\times4.36$ for $1^3\ mm^3$ resolutions.

In images quality comparison, Fig. 4 shows 3 image slices of different reconstructions on 1 GPU and 4 GPUs of both methods, showing identical results over different GPUs implementation. Their horizontal profiles located at the images center in Fig. 5 confirm the same conclusions that reconstructions for both splitting methods are completely identical.

(a) (b) (c)

Fig. 4. Image slices of 3 different reconstructed volumes, for (a) 1 GPU, (b) the volume splitting method on 4 GPUs and (c) the subsets splitting method on 4 GPUs.

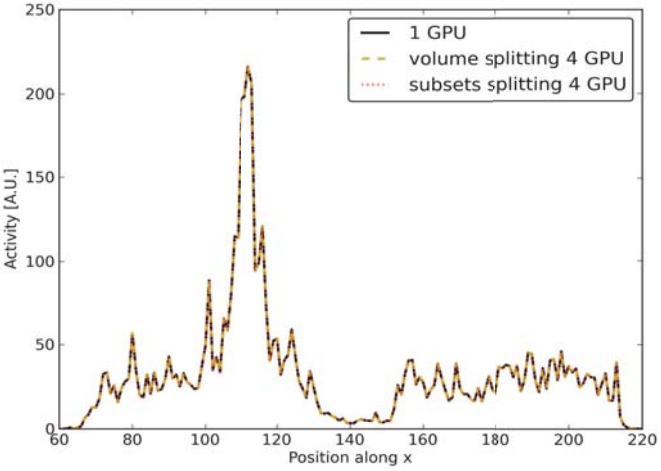


Fig. 5. Profiles through image slices for the 3 different reconstructed volumes on 1 GPU and 4 GPUs for both splitting methods.

IV. Discussion

Regarding to the presented results, we could obviously see the advantages of the volume splitting method over the subsets splitting method on the main obstacles of a fully quantitative PET reconstruction, of performance computing and memory handling. The performance computing issue had been resolved from two parts, minimizing the communication cost and raising the GPU occupancy. In the volume splitting method, the communication part was held on a vector of the correction values that has a small size defined by the LORs number in a subset. This vector size is totally independent of the reconstruction resolution, which explains the very low and constant communication cost. Splitting the volume on portions improves the GPU occupancy by better exploiting the CUDA processing (enough data for CUDA processors). In other part, it resolves the memory handling limitation.

Those two factors of low communication cost and better GPU occupancy enabled a linear scalability performance, which has been actually surpassed with a $\times 4.4$ speedup, where in an ordinal parallelization it will be less than $\times 4$ factor with 4 GPUs use. For the subsets splitting method, the scalability was obstructed with the communication burden, with the speedup decrease on 4 GPUs from $\times 3.88$ in $4^3\ mm^3$ resolution to $\times 3.73$ in $1^3\ mm^3$ resolution, which leads to a saturation status with more GPUs number or higher resolutions.

V. Conclusions

We proposed in this study a new Multi-GPU acceleration method, based on volume parallelization for list-mode PET image reconstruction. Its specificity is optimizing the 3D LM-OSEM algorithm computing time in order to obtain a best run time suitable for a clinical context. We have shown experimental results for the proposed volume parallelization approach with high speedup factors $\times 4.4$ that exceeded the hardware acceleration (4 GPUs) with a very low and stable communication cost over all different resolutions, compared to an acceleration of $\times 3.7$ for the previously proposed solution that was hampered by the high communication cost. The proposed method allows a such acceleration because of the combination of the two factors of the very low communication cost and the improved GPU occupancy. At the same time the memory bottleneck is eliminated by splitting the volume into separate portions on GPUs. With those two points of high performance computing time and memory handling, this work will continue in order to reach such real time for clinical environment with better exploiting our volume splitting method that allows to get a highest Multi-GPUs scalability by multiplying the GPUs number, which is very important for incorporating the corrections needed for a fully quantitative reconstruction.

References

[1] A. Rahmim, J.-C. Cheng, S. Blinder, M.-L. Camborde, and V. Sossi, "Statistical dynamic image reconstruction in state-of-the-art high-resolution pet," *Physics in medicine and biology*, vol. 50, pp. 4887–4912, 2005.

[2] G. Pratx, G. Chinn, P. D. Olcott, and C. S. Levin, "Fast, accurate and shift-varying line projections for iterative reconstruction using the gpu," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 435–445, 2009.

[3] J. Cui, G. Pratx, S. Prevrhal, and C. S. Levin, "Fully 3d list-mode time-of-flight pet image reconstruction on gpus using cuda," *Medical Physics*, vol. 38, pp. 6775–6786, 2011.

[4] J. Bert and D. Visvikis, "A fast cpu/gpu ray projector for fully 3d list-mode pet reconstruction," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 4126 – 4130, 2011.

[5] J. Cui, S. Prevrhal, G. Pratx, L. Shao, and C. S. Levin, "Fully 3-d list-mode positron emission tomography image reconstruction on a multi-gpu cluster," *Fully 3D*, pp. 35–39, 2011.

[6] A. J. Reader, K. Erlandsson, M. A. Flower, and R. J. Ott, "Fast accurate iterative reconstruction for low-statistics positron volume imaging," *Physics in medicine and biology*, vol. 43, pp. 835–846, 1998.

[7] F. Lamare, A. Turzo, Y. Bizais, C. C. L. Rest, and D. Visvikis, "Validation of a monte carlo simulation of the philips allegro/gemini pet systems using gate," *Physics in medicine and biology*, vol. 51, pp. 943–962, 2006.

[8] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, T. Frisson, L. Grevillot, L. Guigues, L. Maigne, C. Morel, Y. P. N. Rehfeld, D. Sarrut, D. R. Schaart, S. Stute, U. Pietrzyk, D. Visvikis, N. Zahra, and I. Buvat, "Gate v6: a major enhancement of the gate simulation platform enabling modelling of ct and radiotherapy," *Physics in medicine and biology*, vol. 56, pp. 881–901, 2011.

[9] J. E. Ortuno, G. Sportelli, P. Guerra, and A. Santos, "Projector model for efficient list-mode reconstruction in pet scanners with parallel planar detectors," *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, pp. 4180–4183, 2011.

[10] G. Sportelli, J. E. Ortuno, and A. Santos, "Efficient rendering of regions of response in list-mode reconstruction for pet," *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, pp. 4219–4221, 2011.

Cone-beam Analytic Reconstruction for Axial Tomography

Jed D. Pack[1], Brian E. Nett[2], Kai Zeng[3], Guangzhi Cao[2], Adam Budde[2], Zhye Yin[1], Bruno De Man[1]
Hye Sun Na[2], Jiahua Fan[2], Kyle Champley[3], Jiang Hsieh[2]

[1] GE Global Research Center, One Research Circle, Niskayuna, NY USA
[2] GE Healthcare, 3000 N. Grandview Blvd., W-1180, Waukesha WI USA
[3] Formerly with GE Global Research Center, One Research Circle, Niskayuna, NY USA

***Abstract*—A comprehensive approach to Cone-beam Analytic Reconstruction for Axial Tomography is proposed where three separate algorithms have been developed to address fundamental reconstruction problems. Cone-beam artifacts are mitigated by addressing each of the root causes separately. The new algorithm is compared to full scan and short-scan FDK algorithms using the anatomically realistic XCAT phantom in both a static and dynamic mode. The new algorithm demonstrates comparable temporal resolution in the cardiac case and for both cardiac and standard imaging significant reduction in cone-beam artifacts is observed.**

I. Introduction

There has been an increase of wide cone-beam systems for: radiation therapy guidance, interventional and surgical guidance, dental CT, breast CT, and micro-CT. Developing accurate and efficient methods for cone-beam image reconstruction from an axial scanning trajectory involves solving three major problems. The first problem is missing frequencies even when acquiring a full 2π scan; it is well-known that there is missing data due to fact that the axial trajectory does not satisfy Tuy's condition [1] for a 3D volume. A second problem (missing frequencies due to z-truncation) arises when one attempts to reconstruct voxels that project outside the detector for some views, or when one uses oblique filtering directions which extend beyond the range of the physical detector. A third problem in axial cone-beam tomography is the handling of redundantly measured data. Frequently, this is done via a view-weighting based on 2D reconstruction algorithms, which can lead to significant mishandling of the cone-beam redundant data. Additional description and references for these issues is given in another abstract for this conference [2].

II. Methods

A. Algorithm Description

As described above there are three fundamental problems to solve in developing an analytic reconstruction method for wide-cone axial tomography. Each of these three problems is addressed with a separate component of the CARAT (Cone-beam Analytic Reconstruction for Axial Tomography) algorithm: Frequency Weighting (FW), Missing Frequency Estimation due to axial Truncation (MFET) and Missing Frequency Estimation (MFE) which occurs throughout the volume (excluding the central slice) due to the axial trajectory. Throughout the manuscript these separate components will be referred to with subscripts.

1) $CARAT_{FW}$: In cone-beam tomography, each measured ray contributes 3D Fourier information in a disk orthogonal to the ray. In order to get an exact reconstruction at a particular voxel, we need to accumulate filtered rays whose total frequency contributions cover the frequency sphere uniformly. With an axial scan, Fourier data is missing at locations outside the scan plane, but we at least wish to have uniform weighting of the data that is available. In the absence of z-truncation, the data at each measured frequency is acquired at two different views (these views correspond to the two views that intersect a Radon plane that corresponds to the selected frequency for the particular voxel). A standard fullscan FDK-type algorithm can be used to average the contributions from these two views together to produce uniform weighting of the measured frequencies with good noise properties. However, there are at least 3 cases for which uniform weighting of these two views is not desirable: i) When temporal resolution is important, we may wish to minimize the contribution of views that are far from a particular selected view, which corresponds to a particular motion phase; ii) In the presence of z truncation, we wish to minimize the impact of a continuous subset of the views at which a voxel of interest projects outside the physical detector (where extrapolated data may be backprojected); iii) When a continuous subset of the views are either missing (e.g., the acquired data is less than a full scan), corrupted, or noisy, we may wish to minimize or eliminate the influence of these views on the reconstructed image.

In all three cases, we desire to keep (or at least overweight) the contribution of the view in each pair which is closer to a particular view, while discarding (or underweighting) the other view in the pair. We have developed an algorithm that allows us to do just this. The process is as follows: First, input data is rebinned from cone beam to cone-parallel geometry, ramp filtered, and cone weighted. Contiguous subsets of axial scan data are backprojected and (in one of two cases) filtered by a Hilbert filter after which they are summed across subsets creating two intermediate image volumes. The image volume generated without the Hilbert filter gives equal weight to each of the two contributions to each frequency component (as does fullscan FDK), while the other contains the difference between the two contributions. A second filter can be applied to the second volume in order to control which of the two contributions receives a positive weight and which receives a negative weight. By summing these two volumes, one

can achieve the desired result of selecting/overweighting the influence of views that are nearer to the center view of the good subset of view data. Cardiac phase selection is controlled by the orientation of this second filter.

The net result is that cardiac reconstructions match or slightly exceed the temporal resolution of short scan reconstructions without the cone-beam artifacts caused by misweighted frequencies. Alternatively, the second volume can be discarded to produce a low noise image volume in the portion of the volume for which 2π radians of data is available. In both cases, the portions of the second image volume where 2π radians of data are not available are filtered by a spatially variant 1D or 2D filter to select and weight only the frequencies that are known to be non-truncated. This novel method creates cardiac reconstructions free from frequency mishandling artifacts and creates both cardiac and standard reconstructions that minimize truncation artifacts and frequency misweighting artifacts.

2) $CARAT_{MFET}$*:* The $CARAT_{MFET}$ algorithm described here acts on the data in the corner regions where less projection data is available. If we examine a given voxel in the far corner, this voxel projects onto the detector for less than π radians of parallel beam data. Thus, the image reconstruction problem faced here is directly analogous to the problem of limited view angle reconstruction or so-called tomosynthesis. Since there is missing data in the corner region due to data truncation there is a desire to estimate this data. If a prior image volume is available such as a helical scan the data could be estimated from that scan. In the case when an additional scan is not available a non-linear filter may be used to reduce the impact of the azmuthial directional streaks that are expected. The aim of this filter is to only correct the data which is problematic in an appropriate conjugate domain. For instance, a 3D x-ray forward projection operator may be used considering a larger virtual detector, or other forward models may be used such as a shift variant Fourier domain model. The fundamental concept is to perform the forward model of both the input image volume f_I and the streak reduced image volume f_{SR}, such that the forward projection is done onto a larger detector in z and the data from the projection of the streak reduced volume is used where real data is not available. In this case the initial reconstruction must be performed on a large enough z grid to support the forward projection operator. A simplified shift variant Fourier model may also be used as the forward and inverse operators. The corrected volume f_C is achieved by applying the inverse operator after the data combination step. In this manner only the ray-sums or frequency components which were corrupt initially are modified during the correction.

3) $CARAT_{MFE}$*:* The $CARAT_{MFE}$ algorithm was developed to solve the missing frequency problem in wide cone axial reconstruction that exists in both short scan and full scan reconstruction. As demonstrated in [3], the missing frequency for a given voxel can be characterized as a cone in the frequency space that contains low in-plane frequency and mid-to-high z-frequency. It is important to note that the missing frequency is shift-variant, i.e., the shape of the cone depends on the voxel location.

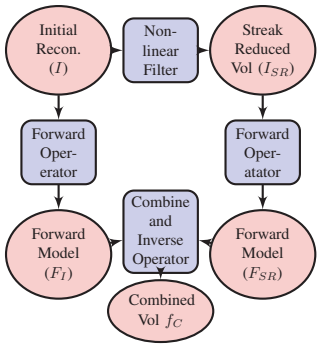


Fig. 1. The schematic diagram of the flow to reduce artifacts from truncated projection data ($CARAT_{MFET}$).

The idea of the $CARAT_{MFE}$ algorithm is to patch in the missing frequencies from a different image volume that contains these missing frequencies. One choice of such an image volume can be helical scout images. If a helical scout is not available, we can also use a similar two-pass approach as proposed in [4] where a first-pass reconstruction is segmented and then used for correction. Specifically, let f be the original wide cone reconstruction that has the missing frequency problem, and f_c the image that contains the missing frequency. First, a filter F is applied on f_c to extract the missing frequencies while removing the complementary frequencies. The filtered result is denoted by $\tilde{f}$. To avoid duplicating frequencies, the complementary filter $\bar{F}$ is applied to the original reconstruction f. Then, we add $\tilde{f}$ back to the filtered original image so that the frequency space is complete in the corrected image $\hat{f}$. The idea is illustrated in Fig. 2(a). It can be seen that the algorithm can be implemented more efficiently as in Fig. 2(b).

As mentioned above, the filter F needs to be shift-variant due to the nature of the missing frequency. Therefore, theoretically for every voxel we need to perform a different filtration. This leads to a significant computation cost. Due to the low in-plane frequency nature of the missing frequencies, both the reconstruction f_c and the frequency patching can be realized on the down-sampled resolution to accelerate the computation. The correction image can then be upsampled (e.g., via interpolation) to the original resolution after filtering.

B. Simulations/ Evaluation

The simulations were performed with the standard male phantom of the XCAT NURBs based phantom, where the forward projection values were generated using the CatSim software package. A cylindrical detector was simulated with a fan angle of 48°, a cone angle of 14.6°, 888 detector channels. The simulations were monoenergetic at 70 keV. In the first case a static phantom was simulated with a beam current of 550 mA and in the second case a dynamic acquisition with a simulated heart rate of 65 bpm was simulated without quantum noise. Comparison with fullscan and short scan FDK were implemented in the cone-parallel geometry and in the fullscan

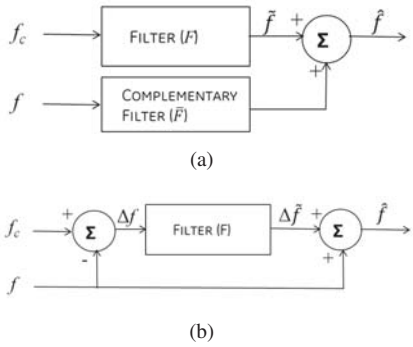


Fig. 2. The algorithm flow of the $CARAT_{MFE}$ method: (a) explicit implementation; (b) improved implementation.

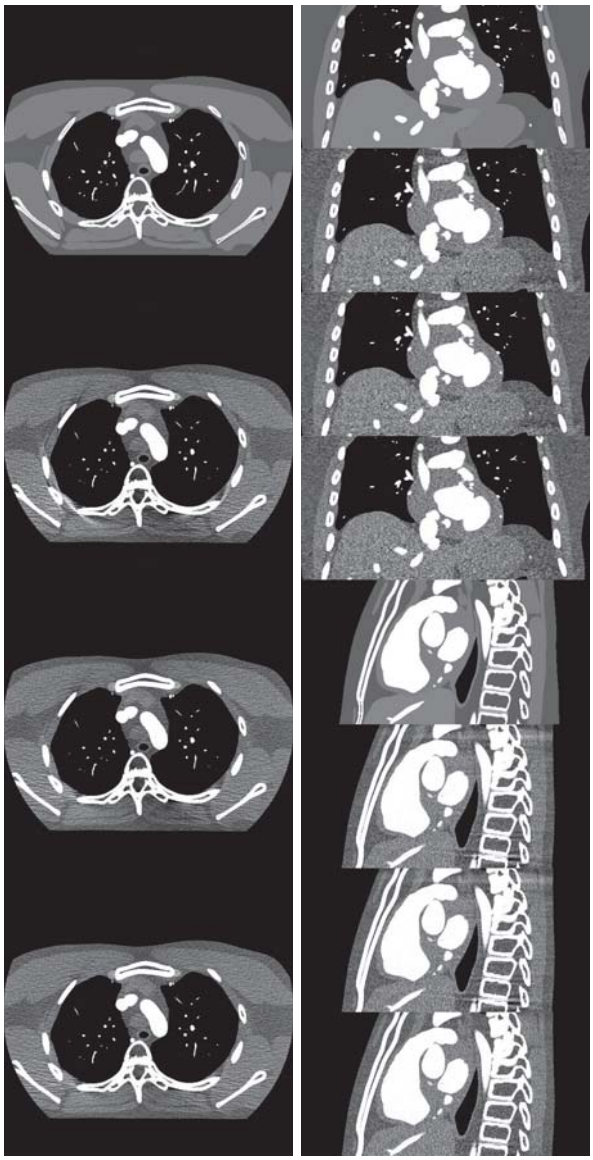
case 3D weighting was used [5]. Images were reconstructed using a centered image volume (in-plane) with a field of view of 320mm, with an image grid on 512×512, where the slices were reconstructed covering a cone angle of $14.6°$.

III. Results

The results will be multiple image comparisons demonstrating the $CARAT$ algorithm including the standard version $CARAT_{STD}$ and the cardiac version $CARAT_{Cardiac}$. Additionally comparisons will be made with standard FDK type reconstructions to demonstrate the improvement in image quality. Due to space constraints detailed quantitative measurements of spatial resolution, temporal resolution, and noise will not be included in this abstract. In Figure 3 images are presented in all three standard reformat planes after each step in the $CARAT$ algorithm with cardiac tuning (i.e. $CARAT_{Cardiac}$). The most obvious improvements in image quality occur in the corner regions after the $CARAT_{MFET}$ algorithm and in the reformat planes after the $CARAT_{MFE}$ algorithm.

After demonstrating the impact of each algorithm within $CARAT$, we show the improvement of the $CARAT_{STD}$ and $CARAT_{Cardiac}$ compared with FDK_{FS} and FDK_{SS} (Figure 4). In comparing the artifacts in the standard reconstructions $CARAT$ has reduced streaks from the ribs in the outer axial slices and improved uniformity in the reformat slices. For cardiac reconstruction the $CARAT_{Cardiac}$ has a significant reduction in cone-beam artifacts compared with FDK_{SS}, this is clearly visible in all planes.

After demonstrating the significant reduction in cone-beam artifacts using the $CARAT_{Cardiac}$ compared with FDK_{SS} we show that in the case of simulated cardiac motion (at 65 bpm, gated in diastole) that the temporal artifacts are comparable between these methods. In Figure 5, it is shown that significant motion artifacts occur when a fullscan of data is used and that the $CARAT_{Cardiac}$ has the desirable property that it has the reduced temporal artifacts as FDK_{SS} without the increased cone-beam artifacts.

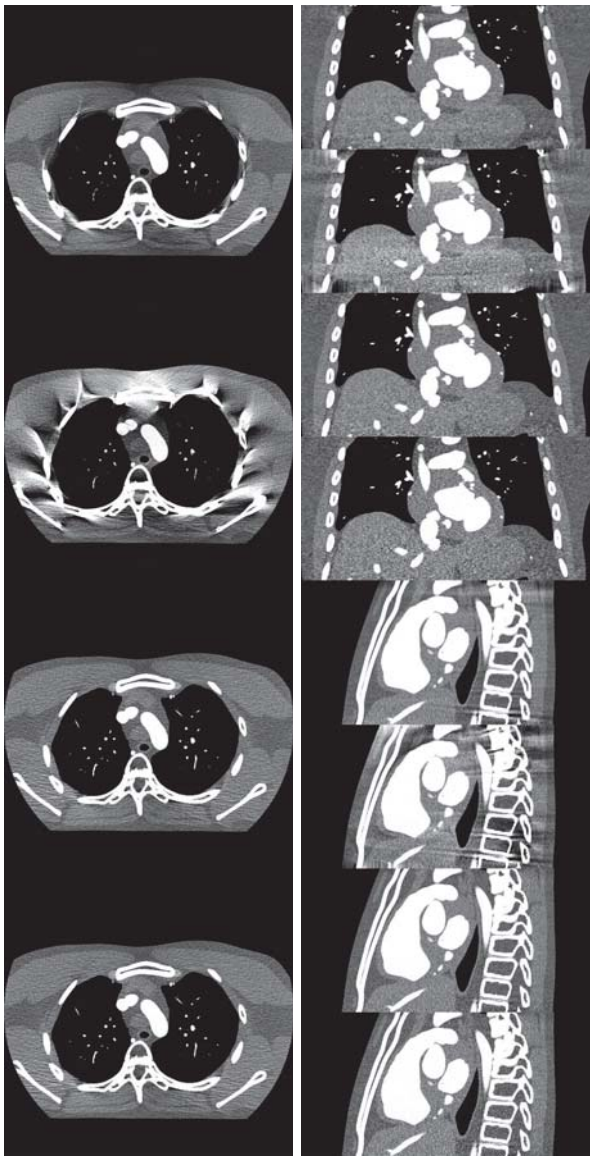


(a) (b)

Fig. 3. Static phantom demonstration of the impact of each step in the $CARAT$ algorithm, in this case with the cardiac tuning. For each panel from top to bottom: Phantom, $CARAT_{FW}$, $CARAT_{FW-MFET}$, $CARAT_{FW-MFET-MFE}$. (a) Axial slice with cone angle of 7.13°, (b-upper) central coronal slice, (b-lower) central sagittal slice. [w/l 400,43] HU. Note in the axial slices the reduction in the streaking from the ribs after $CARAT_{FW-MFET}$ and in the sagittal slices the significant reduction in the streaks from the spine after $CARAT_{FW-MFET-MFE}$

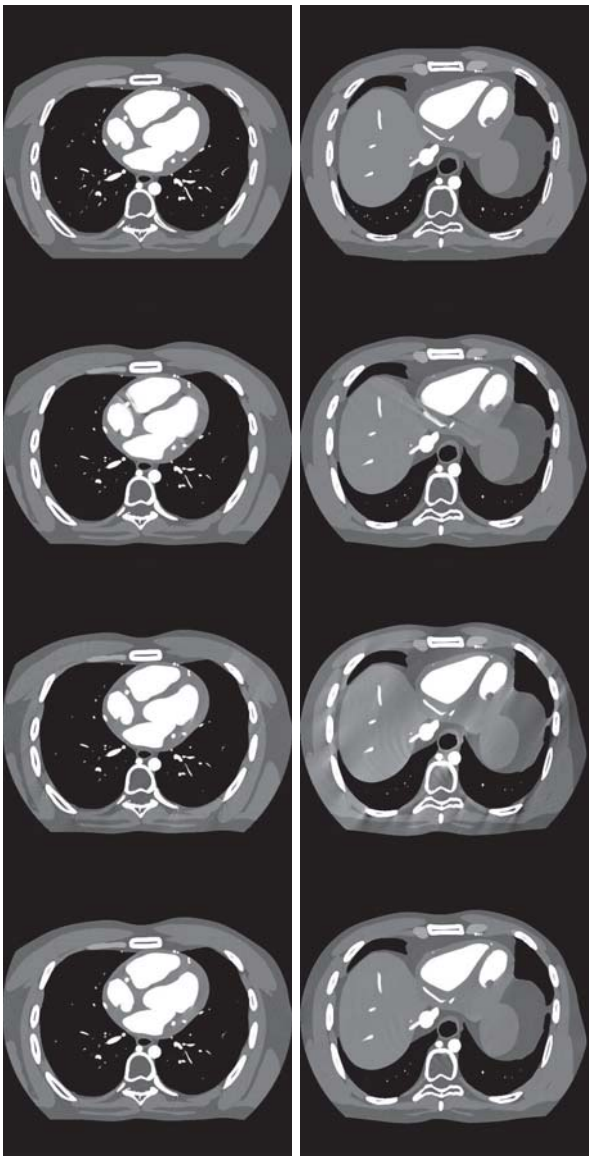
IV. Conclusion

A comprehensive solution has been proposed for Cone-beam Analytic Reconstruction for Axial Tomography ($CARAT$). The algorithm aims at solving the three fundamental problems in widecone analytic image reconstruction: correct frequency weighting, estimation of missing frequency data due to axial truncation and estimation of missing frequency data due to a trajectory lying in a single plane. A multi part framework has been proposed which includes two special cases: $CARAT_{STD}$



(a) (b)

Fig. 4. Static phantom comparison of $CARAT$ with FDK type reconstruction. For each panel from top to bottom: FDK_{FS}, FDK_{SS}, $CARAT_{Cardiac}$,$CARAT_{STD}$. (a) Axial slice with cone angle of 7.13°, (b-upper) central coronal slice, (b-lower) central sagittal slice. [w/l 400,43] HU. (for phantom images see Fig. 4)



(a) (b)

Fig. 5. Simulated 65 bpm cardiac motion. For each panel from top to bottom: Phantom,$CARAT_{STD}$,FDK_{SS}, $CARAT_{Cardiac}$. (a) Axial slice with cone angle of 1.03°, (b) Axial slice with cone angle of 3.31°, [w/l 400,43] HU. The temporal characteristics of the $CARAT_{Cardiac}$ are well matched to the FDK_{SS}, while the $CARAT_{STD}$ has a wider temporal aperture and would not be used for cardiac reconstruction. In (b) the temporal resolution of the FDK_{SS} is sufficient but significant cone-beam artifacts are observed even for the moderate cone-angle of 3.31°.

and $CARAT_{Cardiac}$. Simulations of an anatomically realistic phantom has demonstrated improved image quality for both $CARAT_{STD}$ and $CARAT_{Cardiac}$ compared with the FDK fullscan and shortscan respectively. Initial gated cardiac reconstructions demonstrate that $CARAT_{Cardiac}$ has comparable temporal characteristics to FDK shortscan, but without the significant conebeam artifacts which are present in images reconstructed with shortscan FDK.

REFERENCES

[1] H. K. Tuy, "An inverse formula for cone-beam reconstruction," *SIAM J. Appl. Math*, vol. 43, pp. 546–552, 1983.

[2] J. Pack, Z. Yin, K. Zeng, and B. Nett, "Mitigating cone-beam artifacts in cardiac CT imaging on a large cone-angle system," *Fully 3D Image Reconstruction in Radiology and Nuclear Medicine (Submitted)*, 2013.

[3] S. Bartolac *et al.*, "A local shift-variant fourier model and experimental validation of circular cone-beam computed tomography artifacts," *Med. Phys.*, 2009.

[4] J. Hsieh, "Methods and apparatus for two-pass cone beam image reconstruction," *US patent 6,266,388*, 2001.

[5] X. Tang *et al.*, "A three-dimensional weighted cone beam filtered backprojection (CB-FBP) for image reconstruction in volumetric CT under a circular source trajectory," *Phys. Med. Biol.*, vol. 50, pp. 3889–3905, 2005.

Interior Tomography in a Curvelet Frame

Baodong Liu, Alexander Katsevich, and Hengyong Yu

***Abstract*—Local reconstruction in interior computed tomography (CT) problem is common in practical applications. Both prior knowledge- and compressive sensing-based methods have been developed in recent years to solve the interior problem. A two-dimensional (2D) image can be represented in a curvelet frame. A curvelet, which is localized in both radial and angular directions in the frequency domain, can be used to construct a curvelet frame in the Radon domain. Furthermore, the energy of a curvelet mainly concentrates in a local region. To use these properties, in this paper, we develop a local reconstruction approach for interior problem based on curvelet frames in both image and projection domains. The proposed approach allows a reduction of the size of the system of linear equations that is solved iteratively. Numerical simulation results demonstrate the feasibility of our approach.**

***Index Terms*— Local reconstruction, computed tomography (CT), Radon transform, curvelet transform, biorthogonal curvelet decomposition (BCD).**

I. Introduction

While classic computed tomography (CT) theory targets exact reconstruction of a whole cross-section or of an entire object from complete projections, practical applications often focus on a much smaller internal region-of-interest (ROI), which results in an interior problem. The ROI-focused interior scan entails a reduced radiation dose and can handle larger objects.

The solutions of local reconstruction for interior problem can be roughly categorized into the following generations: (1) For a long time, approximate local reconstruction methods were widely employed because the interior problem does not have a unique solution [1]. (2) The interior problem is exactly and stably solved assuming a known sub-region in the ROI [2-7]. This knowledge-aided solution to the interior problem is named interior tomography. (3) Based on the compressive sensing (CS) theory, an internal ROI is reconstructed uniquely and stably when the ROI is piecewise constant/polynomial [8-10]. Interior tomography reconstruction algorithms can be classified into two types: (1) Iterative reconstruction methods [3, 6-8, 11]. (2) Singular value decomposition (SVD) based differentiated back-projection (DBP) approach [5, 12, 13].

BD Liu and HY Yu are with the Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157 USA. (e-mail: baliu@wakehealth.edu, hengyong-yu@ieee.org). A. Katsevich is with the Department of Mathematics, University of Central Florida, Orlando, FL 32816 USA (e-mail: alexander.katsevich@ucf.edu). Liu and Yu were supported in part by NSF grants CBET- 1149679 and DMS-1210967. Katsevich was supported in part by NSF grants DMS-0806304 and DMS-1211164.

A two-dimensional (2D) function in $L^2(\mathbf{R}^2)$ can be represented by a linear combination of curvelets, which are localized in both radial and angular directions in the frequency domain [14]. The energy of a curvelet mainly concentrates in a local sub-region. For a sub-region outside the ROI, the available interior scan dataset forms a limited angle problem. With a limited scanning angle, if a curvelet's support in the frequency domain is outside the *visible region*, its Radon transform always equals zero [15]. Thus, curvelets with certain orientation localized outside the ROI will be almost invisible to the interior data. Furthermore, a curvelet frame in the spatial domain can be used to construct a "curvelet" frame in the Radon domain. The curvelet coefficients in the spatial domain can be obtained from noise-free Radon data by simply using the frame coefficients in the Radon domain [15, 16].

In this work, we develop an interior tomography approach based on the CS theory and curvelet frames in both image and projection domains. The main idea of the approach is as follows. We split the coefficients of the curvelet expansion of f into three sets. The first set, s_0, consists of the coefficients that can be obtained accurately from the interior data. The second set s_1 consists of the coefficients of f that are essentially invisible from the interior data and can be ignored. All the remaining coefficients need to be determined in an iterative fashion. Since the size of the system that needs to be solved is reduced, this may potentially lead to improved efficiency and noise stability.

II. Method

A. Curvelet Transform

A curvelet frame can be generated by one basic curvelet $\psi_{j,0,0}$ at the scale 2^{-j} ($j \in \mathbf{N}_0$) using dilations, rotations and translations. $\psi_{j,0,0}$ can be defined by its r- and ω-dependence in the polar coordinate system in the frequency domain [14]

$$\hat{\psi}_{j,0,0}(r,\omega) = 2^{-3j/4} W\left(2^{-j} r\right) V\left(2^{\lceil j/2 \rceil + 1} \omega / \pi\right), \qquad (1)$$

where $W(r)$ is a "radial window" and $V(\omega)$ is an "angular window". The windows are smooth, nonnegative and real-valued with $\operatorname{supp} W \subset (1/2, 2)$ and $\operatorname{supp} V \subset (-1, 1)$, and obey the following admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \; r \in (3/4, 3/2);$$

$$\sum_{l=-\infty}^{\infty} V^2\left(\omega - l\right) = 1, \; \omega \in (-1/2, 1/2).$$

The curvelet $\psi_{j,l,k}$ at the scale 2^{-j} with scale-dependent rotation angle $\alpha_{j,l} = l\pi 2^{-\lceil j/2\rceil - 1}$, $-2^{\lceil j/2\rceil+1} \le l \le 2^{\lceil j/2\rceil+1}$ and translation parameters $k = (k_1, k_2)^T \in \mathbf{Z}^2$ can be defined by

$$\psi_{j,l,k} = \psi_{j,0,0}\left(\tilde{R}_{\alpha_{j,l}}\left(x - x_k^{j,l}\right)\right), \tag{2}$$

where $x_k^{j,l} = R_{\alpha_{j,l}}^{-1}\left(k_1 2^{-j}, k_2 2^{-j/2}\right)^T$ is the center of the curvelet, and $\tilde{R}_{\alpha_{j,l}}$ is the rotation matrix defined by $\tilde{R}_\alpha = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$.

Because the region $\bigcup_{j,l,k} \mathrm{supp}(\hat{\psi}_{j,l,k})$ does not cover the whole frequency domain, we need to define a low-pass coarse scale element [14, 15]

$$\hat{\psi}_{-1,0,0}(r,\omega) = W_0(r),\ W_0^2(r) := 1 - \sum_{j=0}^{\infty} W^2(2^{-j}r). \tag{3}$$

The system of curvelets [14, 15]

$$\left\{\psi_{-1,0,k} : k \in \mathbf{Z}^2\right\}$$

$$\bigcup\left\{\psi_{j,l,k} : j \in \mathbf{N}_0, -2^{\lceil j/2\rceil+1} \le l \le 2^{\lceil j/2\rceil+1}, k \in \mathbf{Z}^2\right\}$$

satisfies the tight frame property. An arbitrary function $f \in L^2(\mathbf{R}^2)$ can be represented by a curvelet series

$$f = \sum_n \langle f, \psi_n\rangle \psi_n, \text{ with } n = n(j,l,k), \tag{4}$$

and the Parseval relationship $\sum_n \left|\langle f, \psi_n\rangle\right|^2 = \|f\|^2_{L^2(\mathbf{R}^2)}$ holds. The components in the set $\left\{c_n := \langle f, \psi_n\rangle\right\}$ are called curvelet coefficients.

B. Interior CT Problem

A parallel-beam CT system can be described by the 2D Radon transform of $f \in L^1(\mathbf{R}^2)$ [1]

$$Rf(\theta, s) = R_\theta f(s) = \int_{-\infty}^{\infty} f(s\vec{\theta} + t\vec{\theta}^{\perp})dt, \tag{5}$$

where θ indicates the scanning angle, $\vec{\theta} = (\cos\theta, \sin\theta)$ and $\vec{\theta}^{\perp} = (-\sin\theta, \cos\theta)$. Assuming that f is supported in the unit disk, the interior problem is characterized by the measurements in the range $|s| \le a < 1$, which does not have a unique solution[1]. Let us use R_I to denote the truncated Radon transform data with $|s| \le a < 1$.

We assume that the unknown object f can be represented by a finite linear combination of curvelets [15]

$$f = \sum_{n=1}^{N} \langle f, \psi_n\rangle \psi_n. \tag{6}$$

Let the curvelet coefficient vector be $c := [c_1, c_2, ..., c_N]^T$, where $c_n = \langle f, \psi_n\rangle$, and define operators T and T^* as follows

$$Tf = c,\ T^*c = f. \tag{7}$$

The ideal noise-free interior problem can be expressed as [15]

$$p = Kc, \tag{8}$$

where p represents the measured projection data and $K := R_I T^* = [R\psi_1, R\psi_2, ..., R\psi_N]$.

Our task is to find c knowing the measurements p.

C. Biorthogonal Curvelet Decomposition (BCD)

Based on the curvelet frame $\{\psi_n\}$ and the Radon transform R, we can get systems of functions $\{U_n\}$ and $\{V_n\}$ in $L^2(dsd\theta)$. The systems are generated according to the relationships [16]

$$R\psi_n = 2^{-j} V_n,\ R^* U_n = 2^{-j}\psi_n. \tag{9}$$

The systems $\{U_n\}$ and $\{V_n\}$ are frames and obey the generalized biorthogonality relationships $\langle U_n, V_{n'}\rangle = 2^{j'-j}\langle \psi_n, \psi_{n'}\rangle$. For $\{U_n\}$ we have

$$\left(\sum_n \langle U_n, g\rangle^2\right)^{1/2} \asymp \|g\|_{L^2(dsd\theta)},\ \left\|\sum_n b_n U_n\right\|_2 \le C\left\|(b_n)\right\|_{l^2}.$$

Similar relationships hold for $\{V_n\}$. The BCD of the Radon transform can be expressed with the following reproducing formula [16]

$$f = \sum_{n=1}^{N} 2^j \langle Rf, U_n\rangle \psi_n. \tag{10}$$

This shows that the curvelet coefficient c_n can be obtained from the noise-free Radon data Rf by simply using the $\{U_n\}$ frame coefficient. Therefore, the collection $\{U_n\}$ can be regarded as "curvelets for the sinogram". Given a curvelet ψ_n, the corresponding U_n and V_n can be calculated by [16]

$$U_n = 2^{-j}\left(\square^2 \otimes I\right) R\psi_n,\ V_n = 2^j R\psi_n, \tag{11}$$

where $\left(\square^2 f\right)(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\omega|\, \hat{f}(\omega) e^{i\omega s} d\omega$ and $\left(\square^2 \otimes I\right)$ represents the ramp filter with respect to s for each scanning angle.

D. Algorithm

In the spatial domain, a curvelet is not compactly supported. However, as shown in Fig. 1, its energy mainly concentrates in a small region, whose size is related to the scale of the curvelet.

The functions $U_n(s,\theta)$ and $V_n(s,\theta)$ are smooth on $\mathbf{R}\times[0,2\pi)$ and decay rapidly as $|s| \to \infty$ [16]. Meanwhile, the energy of $U_n(s,\theta)$ also concentrates in a narrow range of s. For those $U_n(s,\theta)$ whose energy is mainly concentrated in the range $[-a, a]$, the following formula holds

$$\langle Rf, U_n\rangle \approx \langle R_I f, U_n\rangle. \tag{12}$$

From (6) and (10), we have

$$f \approx \sum_{n\in S_0} 2^j \langle R_I f, U_n\rangle \psi_n + \sum_{n\notin S_0} c_n \psi_n, \tag{13}$$

where $S_0 = \left\{n \middle| \langle Rf, U_n\rangle \approx \langle R_I f, U_n\rangle\right\}$.

Fig.1. Illustrations of two curvelets at different scale. While the left curvelet has a coarse scale, the right one has a fine scale.

In the Fourier domain, the curvelets are localized with respect to both radial and angular variables. Let l_θ represent a radial line in the polar coordinate system with an angle θ Then we have

$$R_\theta \psi_n \equiv 0 \text{ if } \operatorname{supp}(\hat{\psi}_n) \bigcap l_\theta = \varnothing, \tag{14}$$

which implies that ψ_n is *invisible* for angle θ.

In the interior problem, a point outside the ROI can only be covered by data over a limited angular range. Suppose the energy of a curvelet ψ_n mainly concentrates in a small sub-region Ω_n outside the ROI and Ω_n only can be scanned by the angular range $\theta \in [\theta_1, \theta_2]$. If $\operatorname{supp}(\hat{\psi}_n) \bigcap \{ l_\theta \mid \theta \in [\theta_1, \theta_2] \} = \varnothing$, we get $R_I \psi_n \approx 0$ and the corresponding c_n in (8) can be ignored. Let S_1 denote the index set of this kind of curvelets. Finally, the interior problem (8) can be approximately expressed as

$$p \approx \sum_{n \in S_0} (R_I \psi_n)\left(2^j \langle p, U_n \rangle\right) + \sum_{n \notin S_0 \cup S_1} (R_I \psi_n) c_n . \tag{15}$$

The proposed reconstruction approach for the above problem can be summarized into the following steps:

1) Construct the frame $\{U_n\}$ from a curvelet frame $\{\psi_n\}$ using formula (11).
2) Find the set $S_0 = \{ n \mid \langle Rf, U_n \rangle \approx \langle R_I f, U_n \rangle \}$.

   Because Rf is unknown, the set S_0 can be determined by checking whether the energy of a given U_n is mainly concentrated in the region $[-a, a]$. Let U_n^{ROI} denote the part of U_n located in the region $[-a, a]$ and $h_n := \| U_n^{ROI} \| / \| U_n \|$. Then we define $S_0 := \{ n \mid h_n \geq t_U \}$, where t_U is a threshold.
3) Compute the first vector of curvelet coefficients $c = (c_n)_{n \in S_0}$, where

   $$c_n = 2^j \langle p, U_n \rangle . \tag{16}$$
4) Find $S_1 := \left\{ n \mid \max_{\theta, |s| < a} |(R_I \psi_n)(\theta, s)| < t_C \right\}$, where t_C is a threshold.
5) Set-up a new system of equations

   $$\tilde{p} := p - K_{S_0} c = K_{S_0 \cup S_1} \tilde{c} . \tag{17}$$

   where $K_{S_0} = (R_I \psi_n)_{n \in S_0}$ and $K_{S_0 \cup S_1} = (R_I \psi_n)_{n \notin S_0 \cup S_1}$.
6) Solve equation (17) using an iterative algorithm, e.g. an algebraic reconstruction technique (ART).
7) Get the final image $f = T^*(c^E + \tilde{c}^E)$,

   where $c_n^E = \begin{cases} c_n, & \text{if } n \in S_0, \\ 0, & \text{otherwise} \end{cases}$ and $\tilde{c}_n^E = \begin{cases} \tilde{c}_n, & \text{if } n \notin S_0 \bigcup S_1, \\ 0, & \text{otherwise.} \end{cases}$

Fig. 2. The Shepp-Logan phantom (left) with a display window [0, 1] and the truncated sinogram (right).

III. RESULTS

To verify the feasibility of the proposed local reconstruction method for the interior problem, we implemented it using Matlab and performed numerical simulations. A discrete curvelet transform package (downloaded from *www.curvelet.org*) was used for the curvelet and inverse curvelet transforms.

In our experiments, the imaging object is the Shepp-Logan phantom. The phantom and its projections are generated using the embedded Matlab functions. The phantom is discretized by 64×64 pixels for low computational and memory costs. In each view, the projection is truncated at both ends (1/1.35 of the total data) to simulate the interior problem. The phantom and the corresponding sinogram are shown in Fig. 2.

$\{U_n\}$ is constructed using the discrete curvelet $\{\psi_n\}$. However, we found that the discretized curvelet model and projection calculation method caused serious artifacts in $\{U_n\}$. In our preliminary study, we assume that the curvelet coefficients in S_0 are known and a large threshold t_U =0.9 is used, which results in the number of curvelets in S_0 of about 32.7% of the total number. In step 4), we set t_C =0.02, which eliminates an additional 4.1% of curvelets. Only about 63.2% of curvelets are used in step 5). The reconstructed images are shown in Fig. 3.

To test stability of the proposed method, 3% Gaussian noise is added to the curvelet coefficients in step 3) and the above experiment is repeated. The results are shown in Fig. 4, which confirm the stability of the proposed method.

IV. DISCUSSION AND CONCLUSIONS

From Fig. 3, we can see that the reconstructed image matches well with the original phantom. For the partial result reconstructed using the curvelets in step 3), the accuracy decreases in the sub-regions close to the ROI boundaries.

To make our method practical, in the near future we will continue our research along the following directions. (1) Develop an accurate method to calculate $\{U_n\}$. For each rotation angle θ, $U_n(s, \theta) = \textit{Inverse FT}(\omega\ FT(R_\theta \psi_n))$ According to the central slice theorem, $FT(R_\theta \psi_n)$ is equal to the slice through the origin of the 2D Fourier transform (FT) of

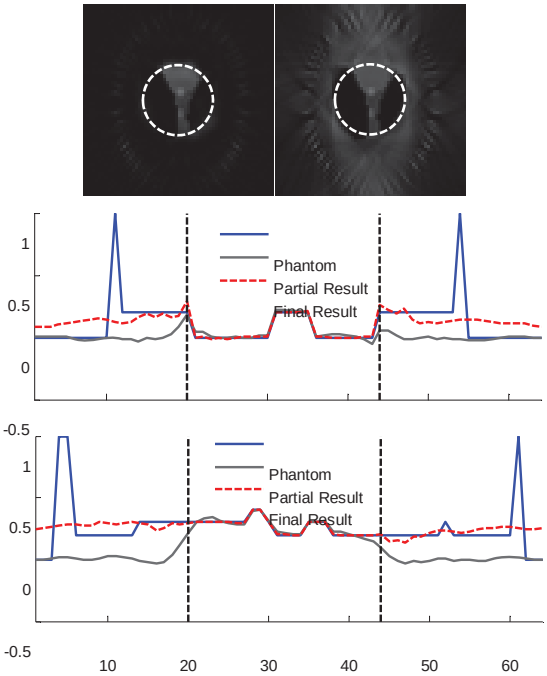


Fig. 3. The results of reconstructions from noise-free projections. The 1st row shows the reconstructed images with a display window [0, 1]. The white dashed circle indicates the ROI. The image on the left is reconstructed using the curvelets in step 3) (called 'partial result'), the image on the right is the final result. The 2nd and 3rd rows show image profiles along the horizontal and vertical central lines, respectively. The ROI region is marked by two black vertical dashed lines.

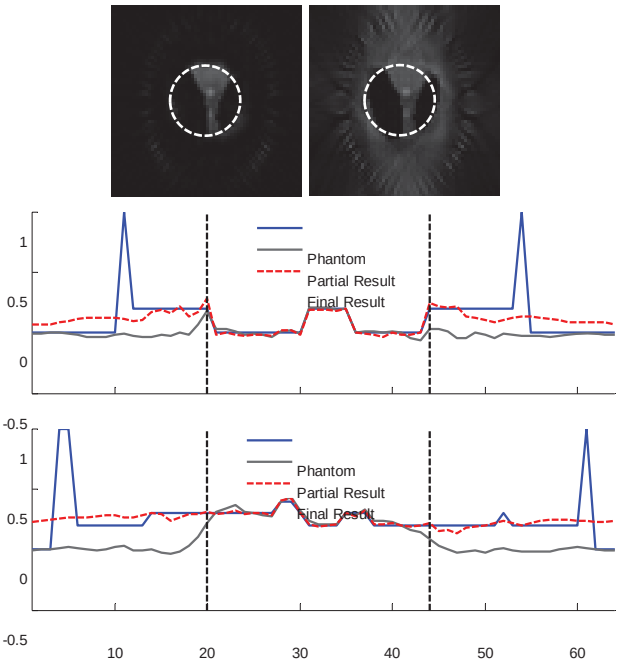


Fig. 4. Same as Fig. 3 but 3% Gaussian noise is added to the curvelet coefficients in step 3).

ψ_n along the direction specified by θ. Therefore, we can compute $FT(R_\theta \psi_n)$ directly from ψ_n using polar FT [17]. Another way we can try is to compute Rf based on $W(r)$ and $V(\omega)$ [15]

$$R_\theta \psi_n(s) = \sqrt{2\pi}\, 2^{j/4} V\left(2^{\lceil j/2 \rceil + 1} \theta' / \pi\right) \cdot \hat{W}\left(2^j \left(\left\langle x_k^{j,l}, (\cos\theta', \sin\theta') \right\rangle - s \right)\right), \quad (18)$$

where $\theta' = \theta + \alpha_{j,l}$. (2) Experiment with more challenging phantoms on a finer grid and real clinical projection datasets. (3) Comparison between the proposed approach and other interior reconstruction methods. (4) Investigation of the relationship between image quality and the thresholds t_U and t_C. (5) Development of the methods to reduce memory and computational costs. We will pursue some of these directions, and the new results will be presented at the conference.

In conclusion, we proposed a new reconstruction scheme for interior CT problem in a curvelet frame. The scheme uses an iterative algorithm for finding only a subset of the curvelet coefficients. A reduction of approximately 37% has been demonstrated. The preliminary experimental results demonstrate the feasibility of the new approach.

REFERENCES

[1] F. Natterer, *The mathematics of computerized tomography*. Philadelphia: Society for Industrial and Applied Mathematics, 2001.

[2] Y. Ye, H. Yu, and G. Wang, "Exact interior reconstruction with cone-beam CT," *International Journal of Biomedical Imaging,* vol. 2007, pp. 1-6, 2007.

[3] Y. Ye, H. Yu, Y. Wei, and G. Wang, "A general local reconstruction approach based on a truncated Hilbert transform," *International Journal of Biomedical Imaging,* vol. 2007, pp. Article ID: 63634, 8 pages, 2007.

[4] Y. Ye, H. Y. Yu, and G. Wang, "Exact interior reconstruction from truncated limited-angle projection data," *International Journal of Biomedical Imaging,* vol. 2008, pp. Article ID: 427989, 6 Pages, 2008.

[5] H. Y. Yu, Y. B. Ye, and G. Wang, "Interior reconstruction using the truncated Hilbert transform via singular value decomposition," *Journal of X-Ray Science and Technology,* vol. 16, pp. 243-251, 2008.

[6] M. Courdurier, F. Noo, M. Defrise, and H. Kudo, "Solving the interior problem of computed tomography using a priori knowledge," *Inverse Problems,* vol. 24, pp. Article ID 065001 , 27 pages, 2008.

[7] H. Kudo, M. Courdurier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Physics in Medicine and Biology,* vol. 53, pp. 2207-2231, 2008.

[8] H. Y. Yu and G. Wang, "Compressed sensing based interior tomography," *Physics in Medicine and Biology,* vol. 54, pp. 2791-2805, May 7 2009.

[9] H. Yu, J. Yang, M. Jiang, and G. Wang, "Supplemental analysis on compressed sensing based interior tomography," *Phys Med Biol,* vol. 54, pp. N425-N432, 2009.

[10] W. Han, H. Yu, and G. Wang, "A total variation minimization theorem for compressed sensing based tomography," *International Journal of Biomedical Imaging,* vol. 2009, pp. Articel ID:125871, 3 pages, 2009.

[11] Q. Xu, X. Q. Mou, G. Wang, J. Sieren, E. A. Hoffman, and H. Y. Yu, "Statistical Interior Tomography," *IEEE Transactions on Medical Imaging,* vol. 30, pp. 1116-1128, May 2011.

[12] A. Katsevich, "Singular value decomposition for the truncated Hilbert transform," *Inverse Problems,* vol. 26, p. 115011, 2010.

[13] X. Jin, A. Katsevich, H. Yu, G. Wang, L. Li, and Z. Chen, "Interior Tomography With Continuous Singular Value Decomposition," *Medical Imaging, IEEE Transactions on,* vol. 31, pp. 2108-2119, 2012.

[14] E. Candes, L. Demanet, D. Donoho, and L. X. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling & Simulation,* vol. 5, pp. 861-899, 2006.

[15] J. Frikel, "Sparse regularization in limited angle tomography," *Applied and Computational Harmonic Analysis,* vol. 34, pp. 117-141, Jan 2013.

[16] E. J. Candes and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Annals of Statistics,* vol. 30, pp. 784-842, Jun 2002.

[17] A. Averbuch, R. R. Coifman, D. L. Donoho, M. Elad, and M. Israeli, "Fast and accurate Polar Fourier transform," *Applied and Computational Harmonic Analysis,* vol. 21, pp. 145-167, Sep 2006.

TV or not TV? That is the Question

Christian Riess, Martin Berger, Haibo Wu, Michael Manhart, Rebecca Fahrig and Andreas Maier

***Abstract*—Iterative reconstruction methods with regularization become more and more popular. In the literature, amazing results are reported that are able to reconstruct images from very few views and from trajectories that do not acquire complete data sets such as would be required for analytical reconstruction methods.**

A large disadvantage of iterative methods is their high computational demand. Bruder et al. have shown that if the reconstruction is accurate enough, the regularization can be performed in the reconstructed image only which allows for much faster application of the regularization term.

In this paper, we present a heuristic compensation weight that corrects for the loss of mass in a filtered back-projection type reconstruction given a limited angle problem. Although the reconstruction contains artifacts, we show that the application of a bilateral filter in the reconstruction domain is able to recover almost the same signal as a TV-regularized iterative reconstruction. The reconstruction error is reduced from 0.130 to 0.057 which is the same as for the iterative case.

I. Introduction

Iterative reconstruction methods that use some kind of regularization are becoming more and more popular. Using certain assumptions, such as that the object of interest is piece-wise constant, allows violation of the Nyquist-Shannon sampling theorem [1], [2]. Reconstruction time, however, is often dramatically increased. Iterative regularized reconstructions are only feasible, if they are implemented on special hardware such as graphics cards. Still, the reconstruction time is an order of magnitude higher than the reconstruction time in a traditional filtered back-projection algorithm.

Recently, novel approaches have been presented providing typical benefits of iterative algorithms, but are based on a filtered back-projection (FBP) type algorithm. Hence, they do not have to project forward and backward repeatedly in their iterations. Bruder et al. have shown that there exists an image-based non-linear filter that is equivalent to a full iterative reconstruction with regularization [3]. However, the method can only be applied if the initial reconstruction is sufficiently accurate. Zeng presented an FBP-type algorithm, which has similar characteristics to those of an iterative MAP (maximum *a posteriori*) algorithm [4].

In general, analytic reconstruction methods face a challenge if they have to reconstruct data from an incomplete trajectory [5], [6]. A fan beam trajectory is complete, if $180^\circ + 2\delta_{\max}$ are acquired, where $\delta_{\max}$ is the half fan angle. This is often referred to as a short scan in the literature. Redundant rays can be weighted which provides a correct reconstruction [7]. Extensions to this weighting to incorporate larger areas of redundancy [8] and to optimize the signal-to-noise-ratio [9] exist. If less than a short scan is acquired, analytic reconstruction is still possible, but the field-of-view (FOV) that allows correct reconstruction is reduced [10]. In contrast, iterative methods using regularization based on total variation (TV) minimization allow the correct reconstruction of the complete FOV, if the object of interest is piece-wise constant [1].

In this paper, we investigate this mismatch. We further propose to use a compensation weight that is a heuristic extension of the commonly used redundancy weights to improve the analytical reconstruction. In order to obtain the final reconstruction, we then apply an image-based regularization using a bilateral filter that enforces piece-wise constancy. Results indicate that this analytical reconstruction delivers reconstructions that are very close to the iterative reconstruction method. Computation time, however, is an order of magnitude lower compared to the iterative procedure.

II. Methods

In the following section, we will shortly describe the used reconstruction methods, beginning with the iterative TV-regularized reconstruction. Next, the analytic reconstruction methods and the different redundancy weights are detailed. At the end of this section, we describe the error metrics that are used in the results section.

A. Iterative Reconstruction

As a reference reconstruction system, we used an iterative reconstruction with an augmented objective function

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

where $\mathbf{x}$ denotes the reconstruction volume, $\mathbf{A}$ the system matrix that projects $\mathbf{x}$ on the detector where the line integrals $\mathbf{b}$ are measured. Details on the implementation are given in [11].

B. Analytic Reconstruction

In the following, we describe the analytical reconstruction algorithm using a 2D formulation. Note that any of the presented concepts can easily be extended to a 3D reconstruction using a Feldkamp-like approximation. The image $f(x, y)$ is reconstructed using a filtered back-projection

$$f(x,y) = \int \frac{1}{U^2} \int \frac{D}{\sqrt{D^2 + s^2}} w(s,\lambda) g(s,\lambda) h_R(s' - s) \mathrm{d}s \mathrm{d}\lambda$$

where D is the distance from the source to the center of rotation, U the depth of the reconstruction point (x, y) and

Christian Riess and Rebecca Fahrig are with the Department of Radiology, Stanford University, Stanford, CA, USA. Martin Berger, Haibo Wu, Michael Manhart and Andreas Maier are with the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative and the Erlangen Graduate School Heterogeneous Imaging Systems.

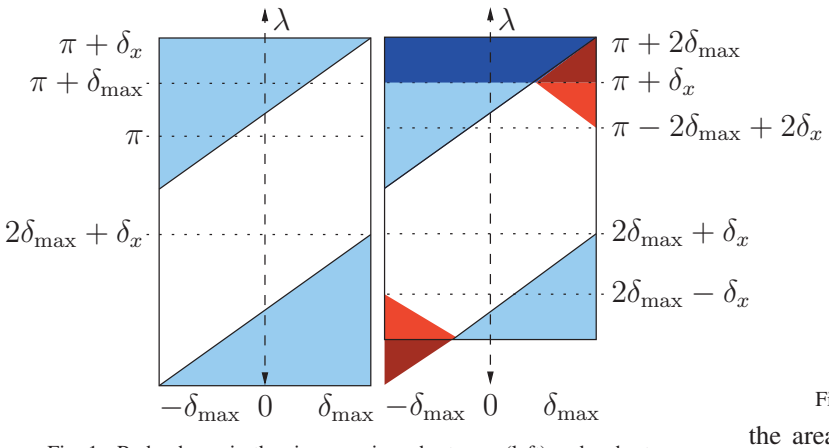


Fig. 1. Redundancy in the sinogram in a short scan (left) and a shorter scan: While the short scan addresses only double rays in the light blue areas, the shorter scan is also missing angles for the complete VOI in the dark red areas.

s' its projection onto the detector $g(s,\lambda)$ at gantry rotation λ. s denotes the detector element and $w(s,\lambda)$ a redundancy weight that deals with inconsistent rays. For convenience, we have virtually placed the detector into the center of rotation.

1) Redundancy Weights: To eliminate artifacts in the reconstruction that are caused by rays which were acquired twice, we use a redundancy weight as described in [8]. Let

$$\eta(\lambda,\delta) = \sin^2\left(\frac{\pi}{2}\frac{\pi+\delta_x-\lambda}{\delta_x-2\delta}\right) \quad \text{and} \tag{1}$$

$$\zeta(\lambda,\delta) = \sin^2\left(\frac{\pi}{2}\frac{\lambda}{\delta_x+2\delta}\right) \tag{2}$$

denote the redundancy weights at the beginning and the end of the scan, respectively. The weights $w_r(s,\lambda)$ are then computed as

$$w_r(s,\lambda) = \begin{cases} \eta(\lambda,\delta) & \text{if} \quad \pi+2\delta \le \lambda \le \pi+\delta_x \\ \zeta(\lambda,\delta) & \text{if} \quad 0 \le \lambda \le 2\delta+\delta_x \\ 1 & \text{otherwise} \end{cases}$$

In this formulation, δ denotes the angle associated with detector element s and δ_x is the scan range in which the redundancy occurs. If $\delta_x = 2\delta_{\max}$ Parker's original formulation is obtained [7]. Note that this formulation is also correct for $\delta_x < 2\delta_{\max}$. The only problem that occurs is that some of the projections ($\lambda = 0$ and $\lambda = \pi+\delta_x$) would get assigned a weight that is 1 for the non-redundant part and 0 for the redundant part. The resulting step function would cause artifacts in the subsequent reconstruction. In order to omit artifacts, we applied a low-pass filter on the weights in these projections.

2) Compensation Weights: Figure 1 shows the difference between a short scan and a scan configuration with $\delta_x < 2\delta_{\max}$. While the short scan only has to solve the redundancy in the triangles described by $\pi+2\delta \le \lambda \le \pi+\delta_x$ and $0 \le \lambda \le 2\delta+\delta_x$ that are shown in light blue in the figure, the shorter scan is missing information in the triangle $\pi+\delta_x \le \lambda \le \pi+2\delta$ that is shown in dark red. The missing data will cause artifacts in the resulting reconstruction. Most of the artifacts are caused by the missing mass during the back-projection. The signal reduction is proportional to the amount of missing angles. The rays in the triangles $\pi+2\delta_x-2\delta \le \lambda \le \pi+\delta_x$ and $0 \le \lambda \le -\delta_x-2\delta$ that are shown in light red pass through

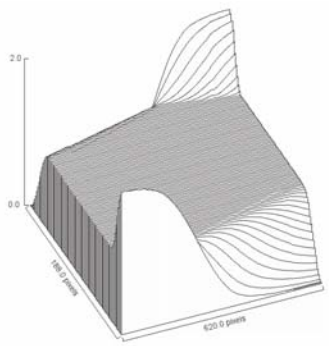


Fig. 2. Surface plot of an instance of compensation weights.

the area where the mass in the reconstruction is missing. In order to create a reconstruction with equal mass distribution, we now increase the weight of these rays with the following compensation weight $w_c(s,\lambda)$:

$$w_c(s,\lambda) = \begin{cases} \eta(\lambda,\delta) & \text{if} \quad \pi+2\delta \le \lambda \le \pi+\delta_x \\ 2-\eta(\lambda,\delta) & \text{if} \quad \pi+2\delta_x-2\delta \le \lambda \le \pi+\delta_x \\ \zeta(\lambda,\delta) & \text{if} \quad 0 \le \lambda \le 2\delta+\delta_x \\ 2-\zeta(\lambda,\delta) & \text{if} \quad 0 \le \lambda \le -\delta_x-2\delta \\ 1 & \text{otherwise} \end{cases}$$

Note that for the projections at $\lambda = 0$ and $\lambda = \pi+\delta_x$ the weight takes the form of a step function as in the case of the redundancy weights. We alleviated the problem by the same low-pass filter as in the case of the redundancy weights. Figure 2 demonstrates the shape of the compensation weights.

3) Bilateral Filtering: The bilateral filter, originally proposed by Tomasi and Manduchi [12], is a smoothing operator that protects sharp edges. The idea is that the spatial support for a Gaussian operator is weighted. More specifically, let $f(x,y)$ and $f^*(x,y)$ denote an intensity in the image at position (x,y) and its bilaterally filtered output, respectively. Here, $f^*(x,y)$ is computed by

$$f^*(x,y) = \sum_{(x',y')\in\mathcal{N}} g(\|(x',y')^{\mathrm{T}} - (x,y)^{\mathrm{T}}\|_2, \sigma_{g,1}) \cdot g(|f(x,y) - f(x',y')|, \sigma_{g,2}) \tag{3}$$

where $g(\mu,\sigma)$ denotes the Gaussian function with mean μ and standard deviation σ, and $\mathcal{N}$ denotes the set of all pixels within a spatially close distance to (x,y).

C. Metrics

For quantitative comparison of the results, we compute three distance metrics, namely the mean square error (MSE), the relative root mean square error (rRMSE) and the structural similarity index (SSID).

The mean square error denotes the pixelwise squared difference between the reconstructed volume and our ground truth, the Shepp-Logan phantom. The relative root mean square error (rRMSE) is similar to the MSE, but normalized with respect to the variations in the image. Thus, it is defined as

$$\epsilon_{\mathrm{rRMSE}} = \frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\tilde{\mathbf{x}}\|} \quad , \tag{4}$$

where $\mathbf{x}$ and $\tilde{\mathbf{x}}$ denote the reconstructed intensities and the ground truth phantom, respectively.

The structural similarity index is a widespread metric that is based on the standard deviation of the reconstructed signal. It is defined as

$$\frac{4 \cdot \mathrm{cov}(\mathbf{x}, \tilde{\mathbf{x}}) \cdot \mu\tilde{\mu}}{(\mu^2 + \tilde{\mu}^2) \cdot (\sigma^2 + \tilde{\sigma}^2)} \quad , \tag{5}$$

where σ and $\tilde{\sigma}$ denote the standard deviations of the reconstructed image $\mathbf{x}$ and the ground truth $\tilde{\mathbf{x}}$, respectively. $\mathrm{cov}(\mathbf{x}, \tilde{\mathbf{x}})$ denotes the covariance between the images. μ and $\tilde{\mu}$ are the mean values of $\mathbf{x}$ and $\tilde{\mathbf{x}}$, respectively.

III. Results

In our experiments, we demonstrate that we achieve a similar image quality using compensation weights and bilateral filtering compared to a state-of-the-art TV-regularized iterative reconstruction [11].

A. Experimental Setup

We evaluated our approach on a simulated 3D Shepp-Logan phantom [13]. For the projection, we used a detector with 640 detector elements of size 0.5 mm and we sampled the phantom at 180 angles with an angular increment of 1. Source to Detector distance was chosen to be 500 mm. The detector was virtually placed into the center of rotation. The phantom was scaled to fill the FOV without truncation As a result of this configuration, the redundancy weighted reconstruction (see next Section) suffers from an undersampled region in the upper part of the image.

B. Qualitative Assessment

Figure 3 shows the qualitative results for the proposed method. On the left, the Shepp-Logan phantom and the result for total variation regularization are shown. In the middle column, the reconstruction results for the classical redundancy weights are shown, with and without bilateral filtering. The right column shows the reconstruction result for the proposed compensation weights.

As expected, Parker weights are not able to reconstruct the area with missing angles correctly. This leads to the dark area in the upper part of the phantom. The TV-regularized reconstruction yields an excellent result. In particular, the sparsity constraint of the algorithm almost perfectly restores the phantom. The reconstruction using compensation weights results in an image with a large number of high frequency streak artifacts that result from the remaining missing information. However, bilateral filtering almost completely removes these artifacts, yielding a result that is comparable to the TV reconstruction result. The same observations can be obtained by looking at the line profiles shown in Figure 4.

C. Quantitive Assessment

For quantitative evaluation, we selected a region of interest (ROI) in the upper part of the phantom, i. e. where the undersampling occurs. Figure 3a shows the region where the ROI was selected.

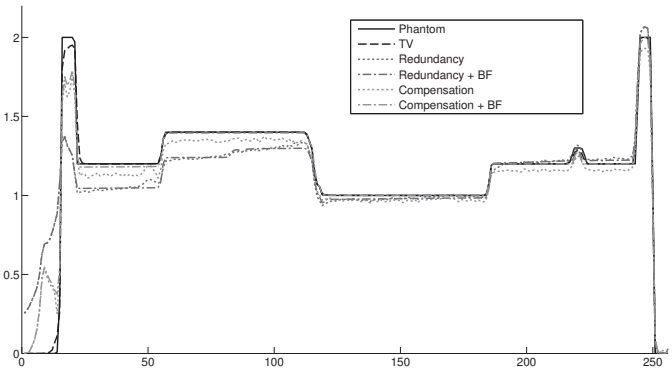


Fig. 4. Intensity profile along a vertical line through the phantom.

	rRMSE	MSE	SSIM
Redundancy	0.1301	0.0286	0.9528
Redundancy BF	0.1271	0.0273	0.9594
Compensation	0.0673	0.0076	0.9594
Compensation BF	0.0569	0.0055	0.9673
TV	0.0566	0.0054	0.9777

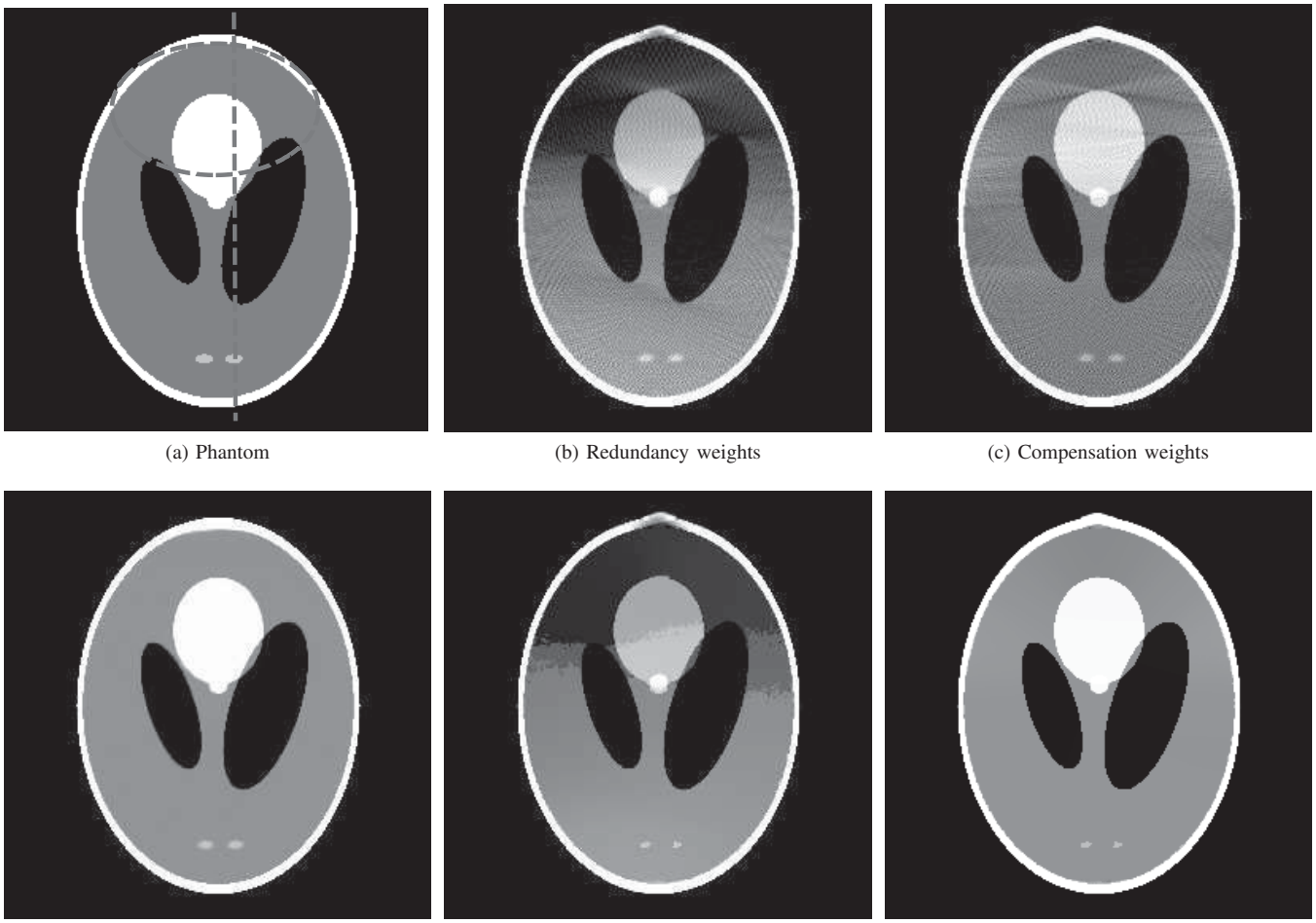
TABLE I
Quantitative measurements.

Table I shows the results of the quantitative measurements. The result of the qualitative assessment is confirmed by all reported measures. The quality of the TV-regularized reconstruction is the best. The reconstruction with compensation weights is very close to this result.

IV. Discussion

Our observations confirm the findings by Bruder et al. [3]. We are able to enforce the regularization by applying a bilateral filter in the reconstruction domain only. This enables us to recover a reconstruction that is comparable to a TV-regularized reconstruction. As commented by Bruder et al. the reconstruction must yield an image quality that is sufficient to enforce the regularization in the reconstruction domain. This is usually not the case in limited angle reconstructions as the missing data leads to a deterministic decrease of reconstruction values in the area with missing angle. We compensate for this using a heuristic weighting procedure that increases weights according to the amount of missing data. Thus we are able to create a reconstruction that is improved but still suffers from streak artifacts. The magnitude of these artifacts, however, is an order of magnitude lower than the artifact resulting from the missing angle. Subsequent use of a bilateral filter is able to recover the original signal.

In the present study, we used a simple phantom that is very popular when exploring reconstructions using TV-regularization. We expect similar results when using other piece-wise constant phantoms [14]. Note that the results presented here required the 8-fold application of the bilateral filter. Still the processing time was an order of magnitude lower than the processing time of the iterative reconstruction with 1000 iterations.



(a) Phantom (b) Redundancy weights (c) Compensation weights

(d) TV-regularized (e) Redundancy weights+BF (f) Compensation weights+BF

Fig. 3. Qualitative comparison of the reconstructed slices: The compensation weights in combination with a bilateral filter (BF) visually appears almost identical to a TV-regularized reconstruction. The window for the visualization was chosen as [1.0, 1.4].

In real data, results may be quite different. However, we still expect that our method is suited to initialize an iterative reconstruction and will therewith decrease the number of required iterations.

V. Conclusion

We have shown a compensation method that allows reconstruction of images that are comparable to the reconstructions created with a TV-regularized iterative reconstruction. The result of the proposed method can be computed within seconds, while the iterative procedure has a computation time that is an order of magnitude higher.

References

[1] E. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[2] X. Pan, E. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse problems*, vol. 25, no. 12, pp. 1–50, 2009. [Online]. Available: http://iopscience.iop.org/0266-5611/25/12/123009

[3] H. Bruder, R. Raupach, J. Sunnegardh, M. Sedlmair, K. Stierstorfer, and T. Flohr, "Adaptive iterative reconstruction," in *SPIE Medical Imaging*, vol. 7961, 2011, pp. 79 610J–79 610J–12. [Online]. Available: http://dx.doi.org/10.1117/12.877953

[4] G. L. Zeng, "View-based noise modeling in the filtered backprojection MAP algorithm," in *Proc. of the 2nd CT Meeting*, 2012, pp. 103–106.

[5] T. M. Buzug, *Computed Tomography*. Berlin, Germany: Springer, 2008.

[6] G. L. Zeng, *Medical Image Reconstruction: A Conceptual Tutorial*. Berlin, Germany: Springer, 2009.

[7] D. L. Parker, "Optimal short scan convolution reconstruction for fan-beam CT," *Medical Physics*, vol. 9, no. 2, pp. 254–257, 1982.

[8] M. D. Silver, "A method for including redundant data in computed tomography," *Medical Physics*, vol. 27, no. 4, pp. 773–774, 2000.

[9] S. Wesarg, M. Ebert, and T. Bortfeld, "Parker weights revisited," *Medical Physics*, vol. 29, no. 3, pp. 372–378, 2002.

[10] F. Noo, M. Defrise, R. Clackdoyle, and H. Kudo, "Image reconstruction from fan-beam projections on less than a short scan," *Physics in Medicine and Biology*, vol. 47, no. 14, pp. 2525–2546, 2002.

[11] H. Wu, C. Rohkohl, and J. Hornegger, "Total Variation Regularization Method for 3-D Rotational Coronary Angiography," in *Bildverarbeitung für die Medizin 2011*, Lübeck, Germany, 2011, pp. 434–438.

[12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998, pp. 839–846.

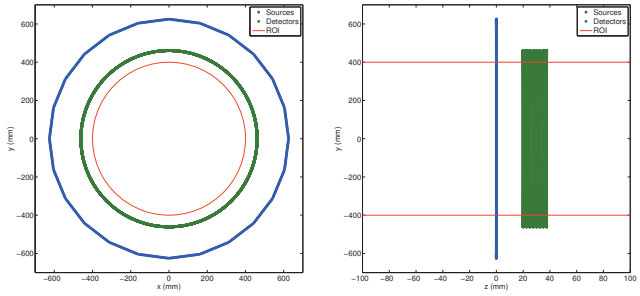
[13] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section-LA Shepp," *IEEE Transactions on Nuclear Science*, vol. NS-21, pp. 21–43, 1974.

[14] A. Maier, H. G. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast simulation of x-ray projections of spline-based surfaces using an append buffer." *Physics in medicine and biology*, vol. 57, no. 19, pp. 6193–210, Oct. 2012. [Online]. Available: http://stacks.iop.org/0031-9155/57/i=19/a=6193

Reduction of Periodic Artefacts for a Switched-Source X-ray CT Machine by Optimising the Source Firing Pattern

William M. Thompson[1], William R. B. Lionheart[1] and Dan Öberg[2]

***Abstract*—The RTT system is a fast cone beam CT scanner which uses a fixed ring of multiple, discrete switchable sources. If the sources are fired in order to approximate a series of helical source trajectories, as in conventional helical cone beam CT, periodic artefacts are observed in the reconstructed images; it is shown that the cause of these artefacts is an insufficient range of illumination angles for certain regions of the reconstruction volume. By adopting an alternative continuum model and optimising the pattern in which the sources are fired to fit this, the range of illumination angles is shown to be much more even over the whole reconstruction volume, leading to the elimination of the periodic artefacts. Examples with simulated and real data are given.**



(a) x-y cross-section (b) y-z cross-section

Fig. 1: The RTT80 geometry

I. Introduction

The Real Time Tomography (RTT) system is a family of fast cone beam CT scanners developed by Rapiscan Systems, designed for applications where scan time is of critical importance. The system achieves its high speed by using a fixed ring of discrete, switchable x-ray sources, and several fixed rings of detectors, offset in the longitudinal direction from the plane of sources. The offset geometry and general concept of the RTT system design are explained in [1].

Specifically, RTT80 is a production RTT system with 80cm tunnel diameter, designed for use in airport baggage screening; the system geometry is shown in figure 1. The system can be operated in either a 'fast' or 'slow' mode, with feed rates of respectively 500 or 250mms^{-1} in each case. In the slow mode discussed in this paper, the complete set of sources is fired so as to simulate a virtual source rotation speed of 15 revolutions per second.

In practice the sources are not fired in strict rotational order, instead being fired such as to approximate a source scanning trajectory of 4 separate intertwined helices. When compared to the longitudinal extent of the detector, the absolute value of the pitch of these helices is large, and this is generally believed to give rise to periodic artefacts in the reconstructed images.

There are two obvious ways to solve this problem; either slow down the feed rate, or increase the longitudinal extension of the detector by adding more detector rings. However, neither of these approaches is desirable, since the feed rate must be kept high enough to avoid creating a backlog of bags on the conveyor feeding the scanner, and adding detector rows would add to the financial cost of the system and also necessitate increasing the detector offset, with negative impact on reconstructed image quality.

The work presented in this paper builds on the authors' previous work in [2], and demonstrates that through careful choice of the order in which the discrete sources are fired, such artefacts can be completely eliminated without using either of the obvious approaches.

1. Henry Moseley X-ray Imaging Facility and School of Mathematics, University of Manchester.
2. Rapiscan Systems.
Corresponding author: William R. B. Lionheart, E-mail: bill.lionheart@manchester.ac.uk.

II. The Firing Order

A secondary advantage of the switched source system design is the ability to switch the sources in almost any order we desire; the order in which this is done is known as a *firing order*, and is defined in general by the periodic extension of the sequence

$$\ldots, \phi(1), \ldots, \phi(N_S), \ldots, \tag{1}$$

determined by some function

$$\phi : \{1, \ldots, N_S\} \to \{1, \ldots, N_S\}, \tag{2}$$

where N_S is the total number of sources in the system. For example, the standard firing order used in the RTT80 system, where $N_S = 768$, is given by the function

$$\phi(i) = \Big[\big(192(i-1) + \lfloor (i-1)/4 \rfloor\big) \bmod 768\Big] + 1, \tag{3}$$

where the operator $\lfloor . \rfloor$ represents the integer part, and is referred to as the *RTT* firing order.

Using conventional CT theory, the firing order is generally chosen to approximate the scanning trajectory that a virtual rotating source or sources would follow. However, it is technically possible to choose the firing order almost

completely arbitrarily; the only limitations are imposed by thermal constraints to allow sources sufficient time to cool after each firing event.

III. The RTT Continuum Model

The standard continuum model for 3D cone beam CT is to consider divergent beam sources located at all points x on some smooth curve $L \subset \mathbb{R}^3$, usually a circle or helix; uniqueness results and analytical inversion formulae for this model are well-known. For conventional 3D cone beam CT, the motivation for choosing this model is clear; physically rotating an x-ray source around a moving object samples source positions on exactly this type of curve.

For switched-source CT systems such as RTT, an alternative continuum model has been proposed, in which divergent beam sources are located at all points x on the surface of a cylinder $C \subset \mathbb{R}^3$. A uniqueness result for this is proved in [3]. The motivation for choosing this model is that in the general case, the firing order may be chosen completely arbitrarily; therefore this can always be chosen so as to generate a regular sampling lattice, and as the number of sources tends to infinity and the time between firing events tends to zero, points on the lattice become arbitrarily close.

IV. An Optimised Firing Order

Adopting the proposed continuum model, it no longer makes sense to think of the firing order as defining the trajectory that a virtual source follows. Instead, we abandon the concept of a source trajectory completely, and regard the firing order as defining an arbitrary sampling pattern of source points on the sampling cylinder C.

Therefore, we seek to choose the firing order so that this sampling pattern is as even as possible. It is shown in [3] that a firing order with this property is given by the function

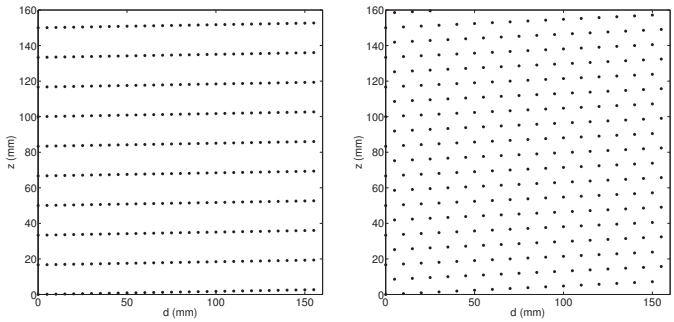
$$\phi(i) = \big[35(i-1) \bmod 768\big] + 1. \tag{4}$$

This firing order is referred to as the $k = 35$ firing order, and also has the important property of *rotational invariance*, in that the sampling pattern created is rotationally symmetric.

V. Reconstruction Algorithms

The standard RTT80 reconstruction algorithm as currently used in production RTT machines is an analytical Feldkamp-type approximate algorithm known as Advanced Cone-beam Back Projection (ACBP) [1]. This is itself a development of the EPBP algorithm [4], modified to cope with the offset detector geometry of the RTT system. The authors are currently working on a similar type of analytical algorithm designed to work with more general firing orders such as $k = 35$.

For the purposes of comparison between the two firing orders, the iterative CGLS algorithm is used [5]; the projection matrix A is calculated using the 3D ray tracing algorithm of Jacobs et al. [6]. In all cases, the reconstruction grid uses cubic voxels of side length $(25/24)$mm.



(a) RTT (b) $k = 35$

Fig. 2: Comparing the sampling pattern on a small section of the sampling cylinder C

VI. Analysis of the Optimised Firing Order

A. The Sampling Pattern

Figure 2 shows a small flattened-out section of the sampling cylinder C, and compares the sampling patterns created by the two firing orders under consideration. Note that the $k = 35$ firing order creates a triangular lattice sampling, with a much more even coverage of the cylinder surface than that created by the RTT firing order.

B. Distribution of Illumination Angles

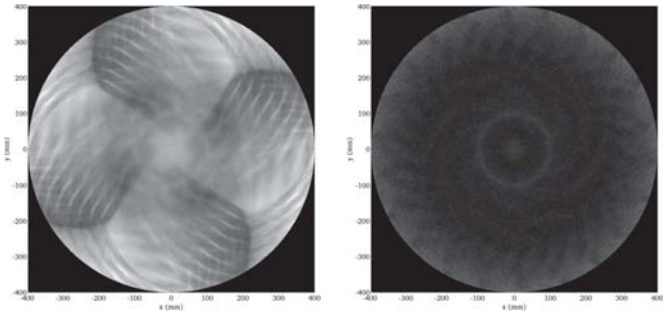
It is possible to study the effect that the firing order is likely to have on reconstructed images by looking at the distribution of angles of illumination across the reconstruction region of interest (ROI). By this, we mean the range of angles $\theta \in [-\pi/2, \pi/2]$ that the rays intersecting a particular region of the ROI make in the x-y plane; for some region R, we refer to this as the *angular distribution* of R.

In the discrete case, the angular distribution of the complete ROI can be studied from the projection matrix A representing all rays intersecting an x-y slice. By the rotational invariance of the firing orders under consideration, the angular distribution of all other slices will be the same up to rotational symmetry.

The projection matrix is calculated on a square grid of cubic voxels, covering the entire ROI. Angles of the rays for all source-detector pairs are calculated, and then ray-voxel intersection lengths are summed in angular bins of a defined width, in this case $10°$. Thus for every voxel, this gives the angular distribution across the defined bins.

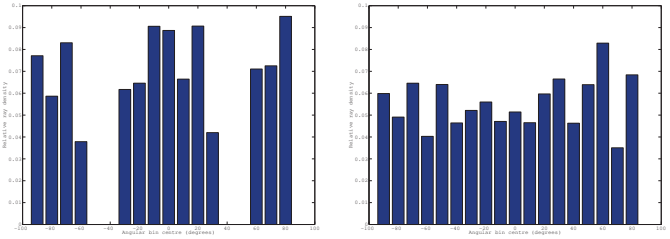
Voxels where the angular distribution is incomplete, or very uneven, will create problems for accurate reconstruction; therefore, as a measure of the evenness of the angular distribution, the standard deviation across the bins is calculated for each voxel. This may be viewed as an image, as in figure 3, which shows that the angular distribution created by the $k = 35$ firing order is in general much more even than that created by the RTT firing order.

In addition to this, the angular distribution for a particular voxel may be plotted; figure 4 shows this for a voxel at the isocentre. Note in particular the distribution for the RTT firing order is incomplete; we therefore expect to see limited angle



(a) RTT (b) $k = 35$

Fig. 3: Images of the standard deviation of the relative ray density across the angular bins for each voxel; grey-scale window in both cases is $[0, 0.05]$



(a) RTT (b) $k = 35$

Fig. 4: Comparing the distribution of relative ray density across the angular bins for a voxel at the isocentre

type artefacts in reconstructions of data obtained with this firing order.

VII. Reconstruction Results

A. Simulated Data

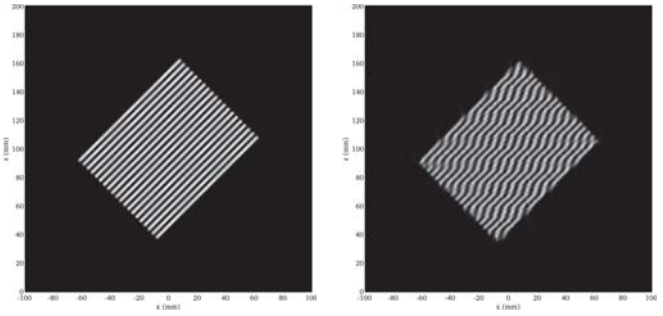
Simulated noise-free data were generated for the Defrise type line-pair phantom shown in figure 5a, consisting of a series of 20 rectangular plates of size 100×10mm, and thickness 2mm, spaced 2mm apart. The phantom is centred at the isocentre and rotated about the y-axis by $45°$.

Figures 5b–5d compare reconstructions by 12 iterations of the iterative CGLS algorithm for both firing orders with reconstruction by the analytical ACBP algorithm for the RTT firing order. As expected, reconstructions from data obtained with the RTT firing order show periodic artefacts, though the exact form of these artefacts depends on the reconstruction algorithm used. However, the reconstruction from data obtained with the $k = 35$ firing order is completely free from these artefacts.

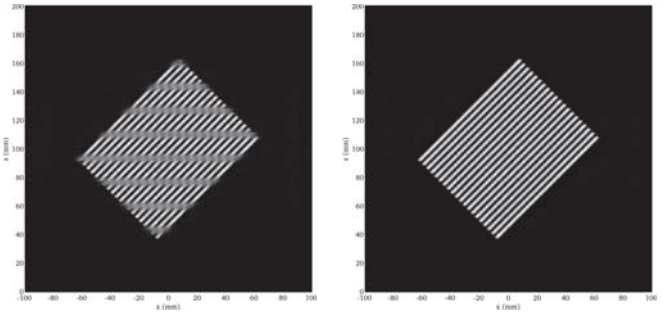
For the CGLS reconstructions, figure 6 shows the 2-norm of the error from the true image at each iteration, defined as

$$\mathbf{e}_i = \|\mathbf{x}_i - \mathbf{x}_{\text{truth}}\|_2, \tag{5}$$

where $\mathbf{x}_i$ is the image at the i^{th} iteration, and $\mathbf{x}_{\text{truth}}$ represents the known true image. Reconstruction from the $k = 35$ data reaches a lower minimum value at 10 iterations, showing a more accurate reconstruction.



(a) True image (b) RTT (ACBP)



(c) RTT (CGLS) (d) $k = 35$ (CGLS)

Fig. 5: Comparing reconstructions of the simulated Defrise type phantom data; grey-scale window in each case is $[0, 1]$ in arbitrary units

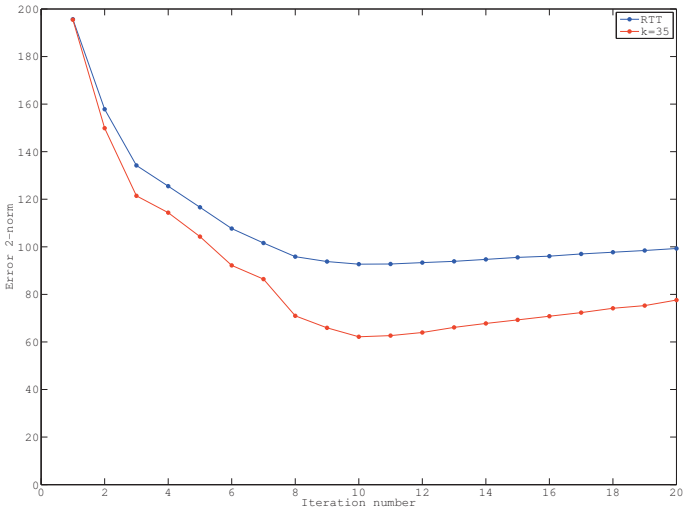


Fig. 6: 2-norm of the image error at each iteration for the simulated data CGLS reconstructions

B. Real Data

Real data were obtained from scans of an ASTM test phantom bag with both firing orders. Figure 7 compares ACBP and CGLS reconstructions of the RTT firing order data with a CGLS reconstruction of the data obtained with the $k = 35$ firing order. Note that the data used for the ACBP reconstruction has been fully pre-processed with corrections for scatter, beam hardening and other artefacts; these corrections were not available for the CGLS reconstructions. In both cases the CGLS reconstructions show the result after 12 iterations.

Reconstructions of the RTT firing order data show periodic artefacts with both reconstruction algorithms, although to a lesser extent in the case of the ACBP algorithm. The CGLS reconstruction of the $k = 35$ data is completely free from such artefacts; note in particular differences on the hinges of the case, at the bottom of the images.

VIII. Conclusion

The work presented demonstrates that by adopting an alternative continuum model for the RTT reconstruction process, and optimising the source firing order to fit this, a more complete data set is obtained. Reconstructions of data collected in this way show complete elimination of periodic artefacts observed in reconstructions of original data sets, in both simulated and real data test cases.

Acknowledgments

This work was jointly funded by EPSRC and Rapiscan, through the University of Manchester Knowledge Transfer Account (project number KTA192).

References

[1] E. Morton, K. Mann, A. Berman, M. Knaup, and M. Kachelriess, "Ultrafast 3D reconstruction for x-ray real-time tomography (RTT)," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, Nov 2009, pp. 4077–4080.

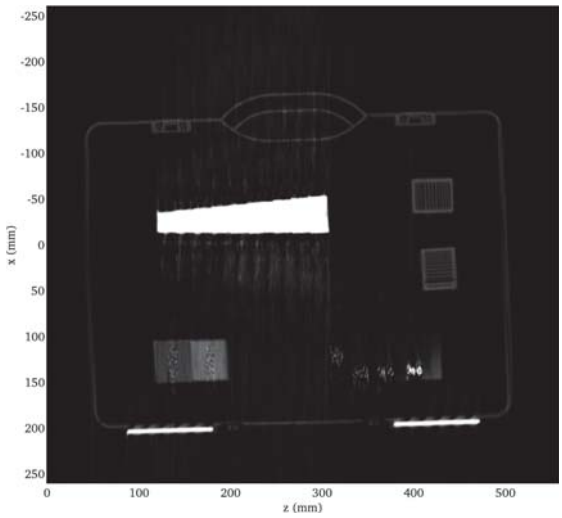
[2] W. M. Thompson and W. R. B. Lionheart, "Optimisation of the source firing pattern for real time cone-beam tomography," in *Fully 3D 2011*, 2011, pp. 335–338.

[3] W. M. Thompson, "Source firing patterns and reconstruction algorithms for a switched source, offset detector CT machine," Ph.D. dissertation, University of Manchester, 2011.

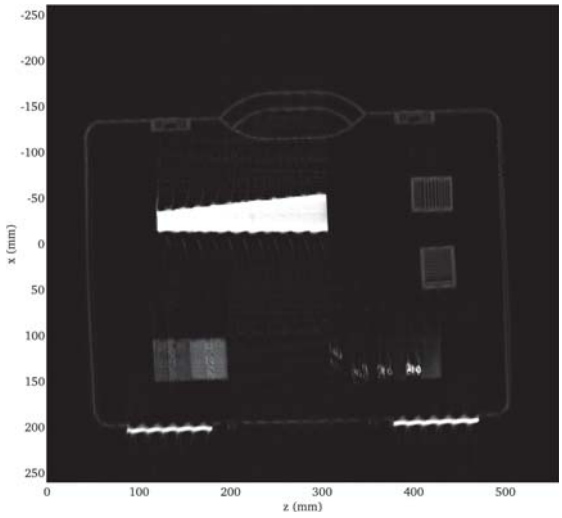
[4] M. Kachelriess, M. Knaup, and W. A. Kalender, "Extended parallel backprojection for standard three-dimensional and phase-correlated four-dimensional axial and spiral cone-beam CT with arbitrary pitch, arbitrary cone-angle, and 100no. 6, pp. 1623–1641, 2004. [Online]. Available: http://link.aip.org/link/?MPH/31/1623/1

[5] P. C. Hansen, "Regularization tools - a MATLAB package for analysis and solution of discrete ill-posed problems - version 3.0 for MATLAB 5.2," *Numer. Algorithms*, vol. 6, pp. 1–35, 1998.

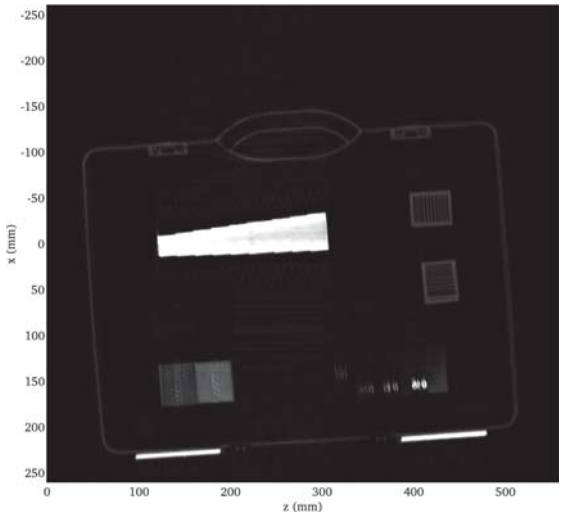
[6] F. Jacobs, E. Sundermann, B. D. Sutter, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *J. Comput. Inform. Technol.*, vol. 6, pp. 89–94, 1998.



(a) RTT (ACBP)



(b) RTT (CGLS)



(c) $k = 35$ (CGLS)

Fig. 7: Comparing reconstructions of the real data; grey-scale window in each case is $[0, 4095]$ in offset Hounsfield units

Fourier-Based Iterative Reconstruction for Limited-Angle Tomography

Bin Yan, Hanming Zhang, Linyuan Wang, Ailong Cai, Xiaoqi Xi and Lei Li

***Abstract*—Image reconstruction from a straight-line trajectory may have interesting applications for industrial or security scanning. Theoretically, the straight-line trajectory violates Tuy's sufficiency condition and suffers from a limited-angle problem. In this article, we propose a Fourier-based iterative approach, based on the recently developed "reconstruction from partial Fourier data"(RecPF) approach, together with a sparse resampling method, to recover the missing information for image reconstruction from a linear scan tomography. The proposed method has some benefits. Firstly, it has computational efficiency as both the forward and back-projection are implemented by the fast Fourier transform(FFT). Secondly, it is efficient by using the classic penalty function approach as well as an alternating minimization technique. Thirdly, it provides improvement in the interpolation accuracy by the sparse resampling strategy. The numerical simulations show that the proposed method brings reasonable performance when applied to limited-angle imaging problem.**

***Index Terms* — straight-line trajectory imaging, limited-angle tomography, Fouried-based iterative reconstruction, sparse resampling.**

I. Introduction

LINEAR scan Computed Tomography (LCT) is of great benefit in the fields of online industrial scanning and security inspection due to its features of straight-line source trajectory and high scanning speed[1]-[4]. However, there are challenges for LCT image reconstruction due to the limited-angle problem. In practical applications of LCT, the length of the detector array could not be infinite. So the projection data are always insufficient for exact reconstruction. And the application of standard analytic algorithms will lead to conspicuous artifacts in reconstructed images. Recently, the combination of real space iterative algorithms with modern optimization methods has been rapidly adopted to deal with the limited-angle problem in CT[5]-[11], while these methods perform well only under certain circumstances, partially due to the presence of insufficient CT data as well as the limit in computation speed. Currently, the most popular method in industrial CT and other tomographic fields remains filtered back projection (FBP) or its variations[12],[13].

Since the Fourier slice theorem provides a Fourier relation for tomographic imaging, a series of iterative Fourier methods[14]-[20] have been exploited and developed, which iterate back and forth between real and Fourier space utilizing the fast Fourier transform(FFT) technique to improve the computational efficiency in tomographic reconstruction. In particular, a Fourier-based iterative algorithm, termed reconstruction from partial Fourier data(RecPF)[21], has been proved to be a powerful technique for recovering magnetic resonance images efficiently and robustly from partial Fourier data.

The purpose of this work is twofold. Firstly, to further extend the capability of RecPF in dealing with the limited-angle tomography problem, we develop and test an iterative Fourier-based tomographic reconstruction structure, based on RecPF approach, in conjunction with a sparse resampling method, to recover the missing information and improve the image quality in LCT reconstruction. We then explore the feasibility of artifact reduction by quantifying the resulting image quality through the numerical simulations of Forbild head phantom. The results show that the proposed method brings reasonable performance when applied to the LCT imaging problem, as measured by the metric of root mean squared error(RMSE). Secondly, we show that the object can be reconstructed from sparse sampling in Fourier domain. This experience should broaden our understanding of the Fourier-based reconstruction issue and provide a new way to reduce the artifacts caused by the interpolations errors in frequency space.

II. Method

A. LCT Scan and Its Associated Sampling in Fourier Space

The configuration of LCT[3] is schematically illustrated in Fig.1(a). Both the x-ray source and the detector would remain stationary and the object scanned is moved on a straight line. This is equivalent to fixing the object and having the source and the detector translated by the object. Fig.1(b) shows the geometry of the LCT scanning. Let O denote the origin in the object coordinates (x, y). An equivalent detector horizontally passing the origin O is used. The source is indexed by l which is its offset from O'. Each detector element is indexed by t which is its offset from O''. The distance from the source to the equivalent detector is D. Hence, the projection $p(l,t)$ is the line integral along the ray AB passing though the source trajectory at l and the detector element at t.

This work was partly supported by the National High Technology Research and Development Program of China (Grant No.2012AA011603)

Authors are with National Digital Switching System Engineering and Technological Research Center, Zhengzhou, 450002, P.R.China. Corresponding author: Bin Yan, E-mail: tom.yan@gmail.com.

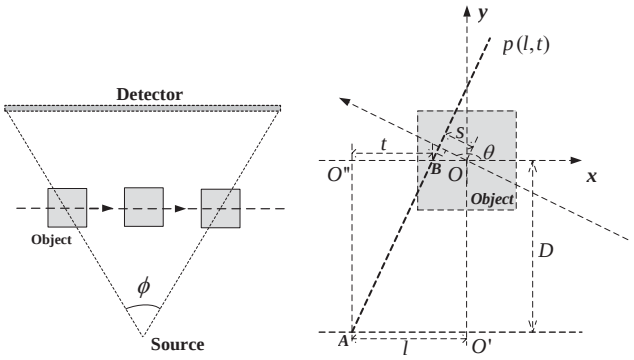


Fig. 1. Systematic description of LCT. (a) The scan geometry; (b) Projection data geometry.

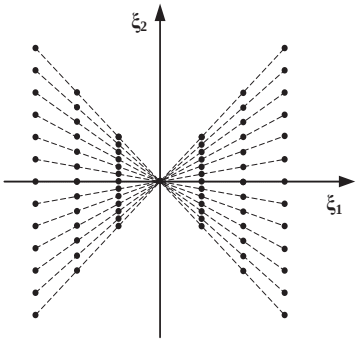


Fig. 2. The sampling points of an LCT scan in the Fourier space.

As in Fig.1(b), one projection data $p(l,t)$ shall correspond to one angle, the data from each detector element corresponds to a certain view angle of parallel beams. Thus, $p(l,t)$ from an object function $f(x,y)$ can be written as

$$\begin{aligned} p(l,t) &= \iint f(x,y)\delta(x\cos\theta + y\sin\theta - s)\ dxdy \\ &= \iint f(x,y)\frac{\sqrt{D^2+t^2}}{D}\delta(x - y\frac{t}{D} - l - t)\ dxdy \end{aligned} \tag{1}$$

with $\theta = \pi - \tan^{-1}(t/D)$ and $s = -D(l+t)/\sqrt{D^2+t^2}$ being the corresponding view angle and detector position in the parallel-beam scan geometry, respectively.

Generally, a part of raw projection data is useless for reconstruction. Data-shift preprocessing $q(l',t) = p(l'-t,t)$ will be useful for saving computation and memory[3]. A Fourier transform of $q(l',t)$ with respect to l' yields

$$\begin{aligned} \hat{q}(\xi,t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{d}l' p(l'-t,t)e^{-j\xi l'} \\ &= \frac{\sqrt{D^2+t^2}}{D\sqrt{2\pi}}\iint \mathrm{d}x\,\mathrm{d}y\, f(x,y)e^{-j\xi(x-yt/D)} \\ &= \sqrt{2\pi}\frac{\sqrt{D^2+t^2}}{D}\hat{f}(\xi, -\frac{\xi t}{D}) \end{aligned} \tag{2}$$

where $\hat{f}$ represents the 2D Fourier transform of f. It can be seen that the sampling of $\hat{q}(\xi,t)$ in Fourier space meets a linogram sampling. The Fourier sampling of projection for the LCT scan is described in Fig.2.

B. Overview of the Reconstruction Scheme

In practical applications, the LCT's effective coverage of projection angle ϕ is less than π. So it is a limited-angle problem and the Fourier sampling is incomplete. To solve this problem, we introduce RecPF method for image recovering from partial Fourier data. The practical reconstruction scheme is performed in three phases. All the steps are summarized below:

- 1D-FFT of the projection to build a linogram 2D Fourier space;
- Linogram to Cartesian resampling;
- Computation of the reconstructed object by RecPF method.

C. Recover Image from Partial Fourier Sampling based on RecPF Method

Let $\bar{u}$ denote the two-dimensional image. And its partial frequency observation is given by

$$f_p = PF\bar{u} + \omega \tag{3}$$

where F represents the Fourier transform matrix, P is a selection matrix and ω represents random noise.

Let $F_p = PF$ and

$$\theta(u, f_p) = (1/2) \bullet \left\| F_p - f_p \right\|_2^2 \tag{4}$$

$\bar{u}$ is reconstructed as a solution of the following model:

$$\arg\min_u \sum_i \left\| D_i u \right\|_2 + \tau \left\| \Psi^{\mathrm{T}} u \right\|_1 + \mu\theta(u, f_p) \tag{5}$$

where $\sum_i \| D_i u \|_2$ is a discretization of the total variation (TV) of u, $\| \Psi^{\mathrm{T}} u \|_1$ is the l_1 norm of the represention u under Ψ, and $\tau, \mu > 0$ are scalars which are used to balance regularization and data fidelity.

By introducing auxiliary variables, $\mathbf{w} = [\mathbf{w}_1, ..., \mathbf{w}_N]$ the augmented Lagrangian method(ALM), which solves each unconstrained subproblem almost exactly, for (5) is an iterative algorithm based on the iteration

$$\begin{cases} (\mathbf{w}, z, u)^{k+1} = \arg\min_{\mathbf{w},z,u} L_A(\mathbf{w}, z, u, (\lambda_1)^k, (\lambda_2)^k) \\ (\lambda_1)_i^{k+1} = (\lambda_1)_i^k - \beta(z_i^{k+1} - \psi_i^{\mathrm{T}} u^{k+1}), \forall i \\ (\lambda_2)_i^{k+1} = (\lambda_2)_i^k - \beta(\mathbf{w}_i^{k+1} - D_i u^{k+1}), \forall i. \end{cases} \tag{6}$$

By utilizing the separable structure of the variables in (5), the cheaper alternating direction method (ADM) is used to decrease L_A at each iteration by just one alternating minimization followed by immediate multiplier updates.

Let $S_1(., \tau/\beta)$ denote the one-dimensional shrinkage operator and $S_2(., 1/\beta)$ denote the two-dimensional shrinkage. The ADM solver implements the following algorithmic framework:

Input: P, f_p, $\tau, \mu > 0$

Algorithm:

Initialize $u = u^0$, $\lambda_1 = (\lambda_1)^0$, $\lambda_2 = (\lambda_2)^0$

Given $\beta > 0$. Set $k = 0$.

While "not converged," **Do**

1) Compute z and w by

$$z_i^{k+1} \leftarrow s_1(\psi_i^{\mathrm{T}} u^k + (\lambda_1)_i^k / \beta, \tau / \beta), \forall i,$$

$$\mathbf{w}_i^{k+1} \leftarrow s_2(D_i u^k + (\lambda_2)_i^k / \beta, 1/\beta), \forall i;$$

2) Compute u^{k+1} by a least squares problem which is diagonalized by a 2-D discrete Fourier transform.

3) Update multipliers λ_1 and λ_2.

4) $k \leftarrow k+1$

End do

Output: u^{k+1}

Fig. 3. Overview of the ADM algorithm..

By the use of alternating minimization scheme, the main computation of the proposed method only involves fast and stable operations consisting of shrinkages and FFTs at each iteration.

D. Sparse Resampling for LCT Fourier Sampling

In the RecPF scheme, the missing data is recovered by the prior constraints and mathematical regularization based on the selection Cartesian points in Fourier domain. Since the Fourier sampling of projections are in concentric squares coordinates and the object is in Cartesian coordinates, interpolation has to be used in the resampling from linogram to Cartesian grid, which introduces artifacts in the reconstructed object. Unlike interpolations in real space where the interpolation error is constrained in the neighboring area, interpolations in Fourier space affect the quality throughout the entire image.

A better solution in this regard is what we call the sparse resampling strategy, which we propose as a new contribution in this paper. Generally, the selection points for RecPF are obtained from the interpolation of linogram sampling in the limited-angle range. The sparse resampling method is to further filter the sampling data in the known angular range: only the Cartesian point which is very close to or coincident with the point of linogram sampling is added to the set of selection points, and the imprecise sampling points are treated as unkown data in the Fourier space. The description of the sparse resampling is shown in Fig.4. The points on orange lines are the sampling of LCT projections, and the green points are the selection points used for a LCT reconstruction.

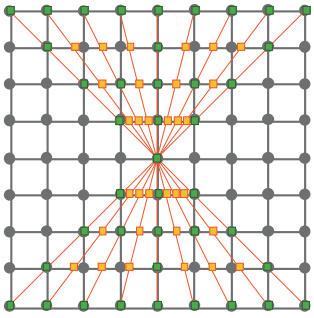


Fig. 4. The description of sparse resampling method in the Fourier space.

In this paper, the above iterative is called "RecPF-SR" (reconstruction from partial Fourier data based on sparse resampling) method. And the reconstruction approach based on RecPF method and b-spline interpolation is termed "RecPF" below.

III. Numerical Experiments

To validate the presented method for image reconstruction, we evaluated it on simulated data of the FORBILD head phantom. In the implementation of RecPF, the parameter of aTV is tuned and gradually changed with interation number for the practical situation.

In a LCT scan, the angular coverage of limited-angle tomography is 130 deg. The perpendicular distances from the source to the object center and to the detector are 240 mm and 480 mm, respectively. The total length of the detector is 2048 mm. The sampling interval in the source direction is 1 mm, and the equivalent interval in the detector direction is 2 mm. The reconstruction has 256×256 pixels with 1mm^2 per pixel size.

In both RecPF and RecPF-SR methods, the oversampling factor in Fourier space is set to 2, so the total number of Cartesian points is $512 \times 512 = 262144$. In the RecPF-SR reconstruction, the number of the sparse resampling points in Fourier space is 13184, the sampling ratio is only nearly 5.03%. The Fourier sampling of RecPF-SR is shown in Fig.5.

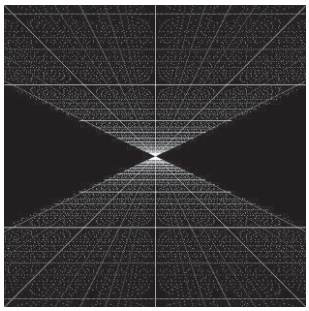
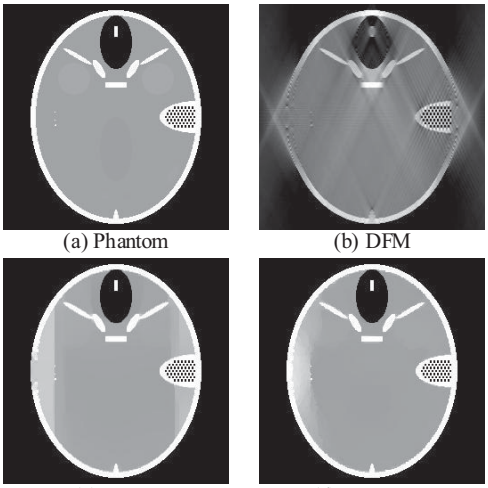


Fig. 5. The sampling points of RecPF-SR method in the Fourier domain.

The images reconstructed from the set of data by using the DFM, RecPF and RecPF-SR algorithms are shown in Fig.6. Each number of iterations for RecPF and RecPF-SR reconstructions is 200. The profiles of these images along the central horizontal rows of the right ear are displayed in Fig.7. It can be seen that the accuracy of the RecPF-SR method is better than the RecPF method.

The convergence speed of these methods are plotted on Fig.8, by evaluating the RMSE after each iteration. We can see that, though RecPF-SR is converged slower in the first, it would obviously get a more accurate reconstruction image after a

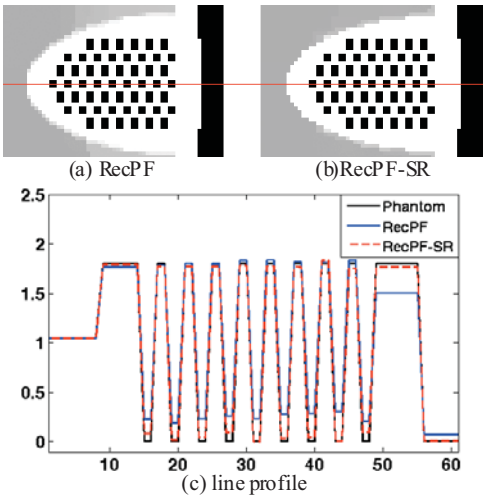
number of iterations.



(a) Phantom (b) DFM

(c) RecPF (d) RecPF-SR

Fig. 6. Image reconstruction of the FORBILD head phantom from a 130° LCT scan. Display window for (a) (c) (d) is [0.85, 1.2],and [0, 1.8] for (b).



(a) RecPF (b)RecPF-SR

(c) line profile

Fig.7.Line profiles along the central horizontal rows of the right ear.

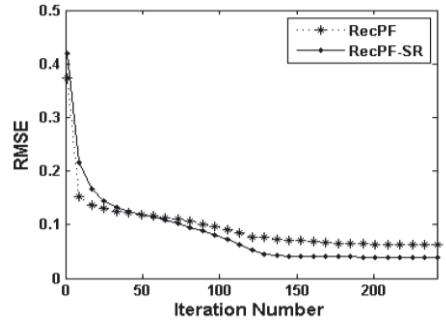


Fig. 8. Root mean squared errors as a function of iterations for RecPF and RecPF-SR methods in a 130° LCT scan.

IV. CONCLUSION

In this paper, We have shown that a RecPF approach could be used to solve limited-angle tomography reconstruction problem. To decline the interpolation error, a sparse resampling method is proposed. Numerical simulations show that both RecPF and RecPF-SR are able to effectively reconstruct high-accuracy images from insufficient sampled data, in particular, the accuracy of RecPF-SR is better than the RecPF method.

REFERENCES

[1] P. Bleuet R. Guillemaud, L. Desbat, and I. Magnin, "An adapted pan volume sampling scheme for 3-D algebraic reconstruction in linear tomosynthesis," *IEEE Trans. Nucl. Sci.,* vol. 49, no. 5, pp. 2366-2362, 2002.

[2] E. Y. Sidky, Y. Zou and X. Pan. "Volume image reconstruction from a straight-line source trajectory," *IEEE Nucl. Sci. Symp. Med. Imag. Conf. Record,* pp. 2441-2444, 2005.

[3] H. Gao, L. Zhang. Z. Chen, Y. Xing, J. Cheng and Z. Qi, "Direct filtered-backprojection-type reconstruction from a straight-line trajectory," *Opt. Eng.* vol. 46 no. 5, pp. 057003, 2007

[4] B. Liu, and L, Zeng. "Parallel SART Algorithm of Linear Scan Cone-Beam CT for Fixed Pipeline," *J. X-ray Sci. and Technol.*, vol. 17, no. 3, pp. 221-232, 2009.

[5] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660-663, 2008.

[6] E. Y. Sidky, and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777-4807, 2008.

[7] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, pp. 1545-1561, 2011.

[8] S. Ramani and J. A. Fessler. "A Splitting-Based Iterative Algorithm for Accelerated Statistical X-Ray CT Reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677-688, 2011.

[9] X. Han, J. Bian, E. Ritman, E. Y. Sidky and X. Pan "Optimization-based reconstruction of sparse images from few-view projections" *Phys. Med. Biol.* vol. 57. pp. 5245–5273, 2012.

[10] L. Wang, L. Li, B. Yan, C. Jiang, H. Wang and S. Bao, "An Algorithm for CT Image Reconstruction from Limited-View Projections," *Chin. Phys. B*, vol. 19, pp. 088106, 2010.

[11] H. Zhang, L. Wang, B. Yan, L. Li, X. Xi and L. Lu, "Image reconstruction based on TV minimization and alternating direction method in linear scan CT," *Chin. Phys. B,* to be published.

[12] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. B: Opt. Phys.*, vol. 1, no. 6, pp. 612-619, 1984.

[13] X. Pan, E. Y. Sidky, M. Vannier, "Why do commercial CT scanners still employ tradition filtered back-projection for image reconstruction?" *Inverse Problems*, vol.25, pp. 123009, 2009.

[14] A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Trans. Image Processing*, vol.5, pp.740-753, 1996.

[15] S. Matej, J. A. Fessler, I. G. Kazantsev, "Iterative tomographic image reconstruction using Fourier-based forward and back-projectors," *IEEE Trans. Med. Imag.*, vol. 23, pp. 401-412. 2004.

[16] Y. Zhang and J. A. Fessler, "Fourier-based forward and back-projectors in iterative fan-beam tomographic image reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, pp. 582-589, 2006.

[17] H. Gao, L. Zhang, Z. Chen, Y. Xing, J. Cheng, and Y. Li, "An extrapolation method for image reconstruction from a straight-line trajectory," *IEEE Nucl. Sci. Symp. Med. Imag. Conf*, pp. 2304-2308, 2006.

[18] H. Gao, Y. Xing, L. Zhang, Z. Chen, and J. Cheng, "Fast and Robust Edge-Preserving Image Reconstruction for Limited-Angle Tomography," *9th Int. Conf. on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine,* pp. 128-131, 2007.

[19] J. Miao, F. Forster, O. Levi, "Equally sloped tomography with oversampling reconstruction," *Phys. Rev. B.*, vol. 72, pp. 052103, 2005.

[20] Y. Mao, B. P. Fahimian, S. Osher, and J. Miao, "Development and Optimization of Regularized Tomographic Reconstruction Algorithms Utilizing Equally-Sloped Tomography," *IEEE Trans Image Process.*, vol. 19, no. 5, 1259-1268, 2010.

[21] J. Yang, Y. Zhang, and W. Yin. "A Fast Alternating Direction Method for TVL1-L2 signal reconstruction from Partial Fourier Data," *IEEE Journal of Selected Topics in Signal Processing*,. vol. 4, no. 2, pp. 288-297, 2010

Simulation of Blurring Artifacts Using a Blur-and-add Model for the Scanning Beam Digital X-ray (SBDX) Tomosynthesis System

Meng Wu, and Rebecca Fahrig

***Abstract*—Tomosynthesis systems with a small tomographic angle have significant out-of-plane blurring artifacts. We propose a tomosynthesis artifact suppression algorithm that utilizes a prior CT volume to augment the run-time image processing. A blur-and-add (BAA) analytical model, derived from the projection-to-backprojection physical model, allows the generation of tomosynthesis images that are a good approximation to the reconstructed image. A computationally practical algorithm is used to simulate images and out-of-plane artifacts from a patient specific prior CT image using the BAA model. The accuracy of the BAA analytical model and the subtraction results were evaluated using patient CT data. Mean-squared-error measurements showed the BAA was accurate when the displacement of volume centers between the BAA and physical model is less than 10 mm. The nodule visibility was improved by subtracting simulated artifacts from the reconstructions, especially in the depth or slice direction.**

***Index Terms*—Tomosynthesis, SBDX, Artifacts suppression**

I. Introduction

Digital tomosynthesis is an imaging technique that has radiation dose levels lower than standard CT, and at the same time, provides depth information that is not available in radiographic fluoroscopy [1]. Through-plane resolution is achieved by providing a relative motion between the x-ray source and the detector through a limited angle about a fulcrum plane, and reconstruction then provides a sharp image of the objects in the plane of interest on which are superimposed blurred images of object details outside the plane of interest.

The most commonly used analytical reconstruction algorithm for tomosynthesis is shift-and-add (SAA). In the case of parallel-path geometry of motion of the tube and/or detector, SAA involves shifting each of the projection images by a given amount and then adding them together. By selecting the shift-per-image amount correctly, objects in a given plane can be brought into sharp focus. In addition to SAA, two deblurring algorithms that have received the most attention in recent years are matrix inversion tomosynthesis (MITS) and filtered backprojection (FBP) [1].

A real-time tomosynthesis approach is available based on the scanning beam digital x-ray (SBDX) hardware, originally developed for cardiac applications by NovaRay, Inc. The SBDX system is an inverse geometry fluoroscopic system with high dose efficiency and the ability to perform continuous real-time tomosynthesis at multiple planes [2], [3]. Due to the very small tomographic angle in the SBDX system, the acquired information from the limited angle distribution is not sufficient to suppress the inter-plane artifacts through reconstruction algorithm only, because the theoretical in-plane and inter-plane resolution ratio is proportional to the tangent of the half tomographic angle [4] which is less than 20 degrees for this system.

In lung nodule transbronchial needle biopsy, patient-specific prior CT volumes are available for biopsy path planning [5]. Those CT images can also be used in tomosynthesis reconstruction or post-processing to suppress out-of-plane blurring artifacts and improve the depth resolution. We want to develop a tomosynthesis reconstruction algorithm that utilizes a prior CT volume to augment the run-time reconstruction from the SBDX system. Yoon et.al. proposed a fast image-based analytical simulation model that produces equivalent tomosynthesis reconstructions to those from a physics-based model [6]. In the analytical blur-and-add (BAA) model, tomographic blurring for the SBDX geometry is modeled as a cylindrical blur function, where the radius of the blur function varies as the distance from the focal plane times the tangent of the half-tomographic angle. Summation through the convolved slices provides an approximation to an SAA tomosynthesis reconstruction of the SBDX system. An important advantage of the BAA model is that each simulated slice is a summation of the blurred planes in the entire volume, so that the blurred out-of-plane structures, tomosynthesis artifacts, can be easily computed by adding specific blurred planes. Subtracting these simulated artifacts can possibly be used to improve the depth resolution.

II. Methods

A. SBDX System

The SBDX system consists of a scanning beam X-ray source with 44-mm-thick 23 cm $\times$ 23 cm collimator and a 10.8 cm $\times$ 5.5 cm CdZnTe detector array, which are mounted on a C-arm gantry [3] [2]. The collimator has an array of 100 $\times$ 100 holes positioned on a 2.3 mm focal spot pitch. We modified the system with 100 cm source to detector distance (D_{sd}) and 60 cm source to central slice (D_{s0}), for which tomographic angles of the region-of-interest (ROI) are 15$\sim$20^o and 5$\sim$15^o in x and y directions, respectively. The conventional SBDX image reconstruction algorithm directly applies shift-and-add to the detector data followed by a pattern correction technique [3]. To use logarithmic data, as in common CT reconstruction,

Meng Wu is with the Electrical Engineering Departments, Stanford University, Stanford, CA, 94305 USA e-mail: mengwu@stanford.edu.

Rebecca Fahrig is with the Radiology Department, School of Medicine, Stanford University, Stanford, CA, 94305 USA e-mail: fahrig@stanford.edu.

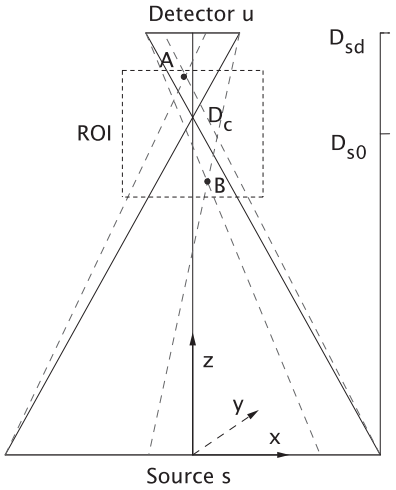


Fig. 1. SBDX system coordinates definition. The space between the source and detector is divided into two region by the critical distance (D_c).

the SBDX system needs to have more photon counts per pixel. Therefore we proposed to use more repeated scan data as well as binning the detector to 48×24 with 2.28×2.28 mm^2 pixel size.

The system geometry and coordinates are shown in Fig. 1 in two dimensions. The source lies on points on line s parallel to the x axis. The flat detector pixels lie on the line u parallel to the x axis. The planes of the reconstruction volume are parallel to both source and detector planes. The center of the source plane and the center of the detector plane define the vertical z axis. The origin of z axis is defined at the source plane. Since this system is separable in x and y directions, we described our method in the x-z plane for simplicity.

B. Blur-and-add Model

Assuming the centers of the detector and source planes are perfectly aligned, lines connect opposite boundaries of the detector and source plane intersect at:

$$D_c = \frac{l_{src}}{l_{det} + l_{src}} D_{sd}, \tag{1}$$

where l_{det} and l_{src} are the widths of detector and source plane, respectively. This critical distance (D_c) divides the space into two regions. Most of the voxels above D_c (point A in Fig. 1) can "see" the entire source plane but only part of detector. The x-ray, passing through this region, is mainly truncated by the finite size of source plane. In the same way, most of the voxels below D_c (point B in Fig. 1) can "see" the entire detector plane but part of the source plane. The SAA back projected image is:

$$\begin{aligned} f_b(x,z) =& \int_s \int_u \delta(x - \frac{uz}{D_{sd}} - \frac{s(D_{sd}-z)}{D_{sd}}) \int_{z'} \int_{x'} \\ & f(x',z')\delta(x' - \frac{uz'}{D_{sd}} - \frac{s(D_{sd}-z')}{D_{sd}}) \mathrm{d}x'\mathrm{d}z'\mathrm{d}u\mathrm{d}s \\ =& f_b^{\text{lower}}(x,z) + f_b^{\text{upper}}(x,z). \end{aligned} \tag{2}$$

where $f(x,z)$ denotes the image voxel values at the 2D spatial location (x,z), and $f_b^{\text{upper}}(x,z)$ and $f_b^{\text{lower}}(x,z)$ denote the content of the SAA reconstructed image contributed by the region above and below the critical distance. The delta functions imply the point-to-point projection process in the SBDX system. We temporarily ignored the ray weighting factors in both forward and backward projections because they can be canceled in projectors.

The x-ray, passing through the lower region, is mainly truncated by the finite detector. For the lower region, the backprojected image is

$$\begin{aligned} & f_b^{\text{lower}}(x,z) \\ \approx & \int_0^{D_c} \int_u \int_{x'} f(x',z')\delta(x' - x\frac{D_{sd}-z'}{D_{sd}-z} + u\frac{z-z'}{D_{sd}-z}) \\ & \frac{D_{sd}}{D_{sd}-z} \cdot \mathrm{rect}(\frac{x\frac{D_{sd}}{D_{sd}-z} - u\frac{z}{D_{sd}-z}}{l_{src}}) \mathrm{d}x'\mathrm{d}u\mathrm{d}z' \\ = & \int_0^{D_c} \int_{-l_{det}/2}^{+l_{det}/2} \frac{D_{sd}}{D_{sd}-z} f(x\frac{D_{sd}-z'}{D_{sd}-z} - u\frac{z-z'}{D_{sd}-z}, z') \\ & \mathrm{rect}(\frac{x\frac{D_{sd}}{D_{sd}-z} - u\frac{z}{D_{sd}-z}}{l_{src}}) \mathrm{d}u\mathrm{d}z'. \end{aligned} \tag{3}$$

The term $x\frac{D_{sd}-z'}{D_{sd}-z} - u\frac{z-z'}{D_{sd}-z}$ describes the geometric deformation of the SBDX system. The rectangular window function denotes truncation by the finite size of the source plane. We moved the truncation function into the blurring function by approximating truncation at the central ray. Then the shift-invariant BAA model for the lower region is

$$\begin{aligned} f_b^{\text{lower}}(x,z) \approx & \int_0^{D_c} \frac{a(D_{sd}-z)}{(D_{sd}-z')|z'-z|} \mathrm{rect}(\frac{x\frac{D_{sd}-z'}{|z'-z|}}{l_{det}}) \\ & * \left[f(x\frac{D_{sd}-z'}{D_{sd}-z}, z') \mathrm{rect}(\frac{x\frac{D_{sd}}{D_{sd}-z}}{l_{src} + l_{det}\frac{z}{D_{sd}-z}}) \right] \mathrm{d}z'. \end{aligned} \tag{4}$$

For the lower region, the truncation is dominated by the detector, therefore we ignored the window function due to the source plane. Within certain field-of-view we could the ignore the rectangular window function caused by the source plane. The blurring kernel is a function of plane-to-source distance z and plane-to-plane distance $z'-z$. Then we obtain the convolution-based BAA model

$$\begin{aligned} f_b^{\text{lower}}(x,z) \approx & \int_0^{D_c} \frac{a(D_{sd}-z)}{(D_{sd}-z')|z'-z|} \\ & f(x\frac{D_{sd}-z'}{D_{sd}-z}, z') * \mathrm{rect}(\frac{x\frac{D_{sd}-z'}{|z'-z|}}{l_{det}}) \mathrm{d}z'. \end{aligned} \tag{5}$$

In the same way, the BAA model for the upper region is given by:

$$f_b^{\text{upper}}(x,z) \approx \int_{D_c}^{D_{sd}} \frac{az}{z'|z'-z|} f(x\frac{z'}{z}, z') * \mathrm{rect}(\frac{x\frac{z'}{|z'-z|}}{l_{src}}) \mathrm{d}z'. \tag{6}$$

C. Blur-and-add Artifact Simulation

To simulate SBDX tomosynthesis reconstruction analytically from CT images using the BAA model, we directly

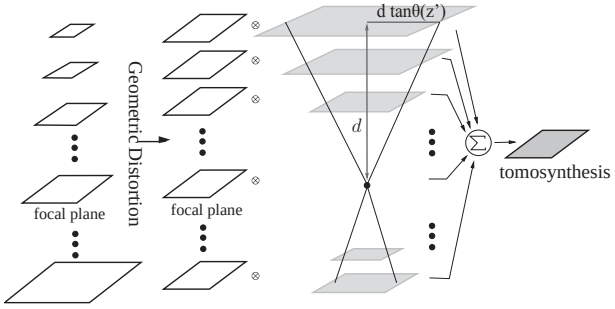


Fig. 2. Schematic drawing of the proposed fast convolution-based BAA model with rectangular blurring kernel for SBDX system . Voxels on each plane are first geometrically distorted following the center lines determined by the system geometry and the plane z location. Geometrically distorted planes are then convolved with an appropriately normalized blurring kernel determined by the system geometry and the source-detector pair trajectory. To simulate a tomosynthetic reconstruction onto a specified focal plane, all the convolved plane data are summed in the z direction.

applied blurring to the volume data on a plane-by-plane basis and summed the blurred planar images into a single tomosynthesis image plane as

$$f_{\text{sim}}(x,z) \approx \sum_{z'=0}^{D_{sd}} a(z,z')\left[f(x\cdot d(z,z'),z') * h(x,z;z')\right], \tag{7}$$

where $h(x,z;z')$ is the blurring kernel of the plane at z' relative to the tomosynthesis plane at z. The blurring kernel is combined rectangular functions in (5) and (6). $a(z,z')$ is the amplitude function such that

$$a(z,z') = \begin{cases} \dfrac{\Delta_z(D_{sd}-z)}{(D_{sd}-z')} & \text{if } z' < D_c \\ \dfrac{\Delta_z z}{z'} & \text{if } z' > D_c; \end{cases} \tag{8}$$

and $d(z,z')$ is the geometric distortion function such that

$$d(z,z') = \begin{cases} \dfrac{D_{sd}-z'}{D_{sd}-z} & \text{if } z' < D_c \\ \dfrac{z'}{z} & \text{if } z' > D_c. \end{cases} \tag{9}$$

We used the blurring kernel at the center of each plane as an approximation to the blur over the entire plane. This approximation is quite accurate when the voxel is close to the z axis. We could then use discrete convolution or FFT to compute the BAA simulated image. Putting all the components together, the BAA model is illustrated in Fig.2.

The ultimate goal of this study is to use a prior CT volume to suppress out-of-plane structures that dominate the artifacts in the tomosynthesis images. Since the blur-and-add model describes tomosynthesis images on a plane-to-plane basis, it is also easy to compute the undesired out-of-plane structures within each reconstructed plane. The BAA-simulated tomosynthesis artifact is given by

$$f_{\text{art}}(x,z;t) = \sum_{|z'-z|>t\Delta_z} a(z,z')\left[f(x\cdot d(z,z'),z') * h(x,z;z')\right], \tag{10}$$

where Δ_z denotes the slice thickness of images and t denotes the number of slices that will be preserved. The out-of-plane

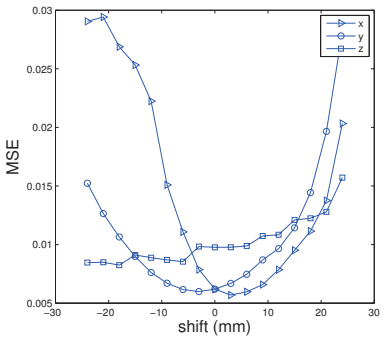


Fig. 3. Accuracy assessment of the BAA analytical model against shift of volume in x, y and z direction.

structures from the neighboring planes within the preserved thickness is not considered to be artifacts $f_{\text{art}}(x,z;t)$. Moreover, instead of using a sharp step-like function to define the preserved slice thickness, we suggest using a continuous function such as

$$\begin{aligned} f_{\text{art}}(x,z;k) = &\sum_{z'=0}^{D_{sd}} (1-\exp(-\frac{|z-z'|}{k\Delta_z})) \\ &\times a(z,z')[f(x\cdot d(z,z'),z') * h(x,z;z')], \end{aligned} \tag{11}$$

where the value of the suppression parameter k controls how aggressively we want to suppress the tomosynthesis artifacts. Assuming that we have accurate 3D registration results between the SAA reconstructed image and the BAA simulated image, the inter-plane artifact in the tomosynthesis image can be suppressed by subtracting simulated artifacts from the SAA reconstruction using

$$f_{\text{sub}}(x,z;k) = \frac{f_{\text{recon}}(x,z)-\mu_{\text{recon}}}{\sigma_{\text{recon}}} - \frac{f_{\text{art}}(x,z;k)-\mu_{\text{sim}}}{\sigma_{\text{sim}}}, \tag{12}$$

where $f_{\text{art}}(x,z;k)$ is the simulated artifact using the BAA model. Both images are normalized before the subtraction if they are at different magnitude levels. μ and σ denote the mean and deviation of the images within the reconstructed volume.

III. Simulation and Results

We used the real lung cancer patient CT scan that includes a pulmonary nodule of 12 mm in diameter. The CT image size was 512×277×512 with spacing of 0.82×1.25×0.82 mm^3. The CT image was shifted to place the target nodules at the center of the ROI. The reconstruction volume size was 128×128×32 with spacing of 0.5×0.5×3 mm^3 in the x, y, and z directions, respectively. The simulated image using the BAA analytical model had a volume size of 256×256×52 with the same spacing as the physical model.

To demonstrate the feasibility of the BAA analytical model, we compared images using both the SAA physical model and the BAA analytical model from the same CT scans. We used ray-driven and pixel-driven methods for projection and backprojection for (SAA) in the physical model. No additional noise was added to the simulated projections or reconstructed images. The BAA analytical model approximates the blurring kernel of each plane by the kernel corresponding to that of the voxel on the central axis. Therefore, the accuracy of the

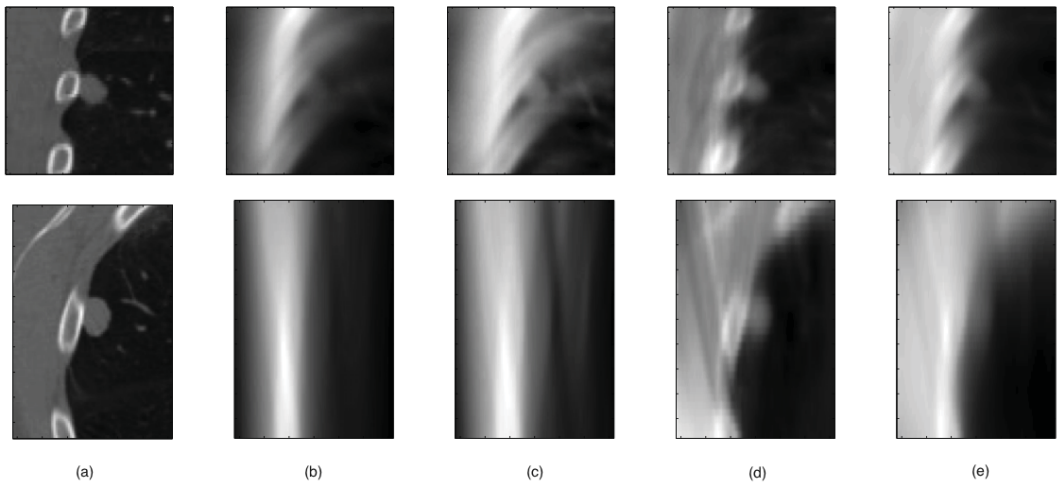


Fig. 4. Subtraction-based artifact suppression images. The columns from left to right are: (a) CT images, (b) SAA reconstructions, (c) ML reconstructions, (d) subtraction images with k = 2, (e) subtraction images with k = 8. The last column shows the simulated projection images. The top row is the coronal view, and the bottom row is the axial view. The images are scaled to full range of colormap by Matlab.

analytical model may decrease as the reconstruction center moves away from the simulation center. To evaluate the errors caused by volume displacements, we simulated reconstruction images with shifts from -25 to +25 mm in x, y and z directions separately. Mean-square-errors (MSE) were then calculated between the normalized (mean zeros and unit variance in ROI) reconstructed images and its corresponding volume in BAA simulated image.

Fig. 3 shows the accuracy of the BAA analytical model at different displacements of patient CT data sets. The accuracy of the analytical model decreases as the center of the reconstruction volume moves away from the central axis in the x and y directions. The displacement in z has less influence on the accuracy, because it has small effect on the tomographic angles of voxels and on geometric distortions. According to MSE measurements, there is a range of shifts from -10 mm to +10 mm in x and y directions, for which displacement does not greatly influence the accuracy of the analytical model. Therefore, we set 10 mm as the range for which the BAA analytical model is accurate. Displacements larger than this require either moving the patient or re-simulating the image. The range gives us a 20×20 mm^2 tolerable motion of reconstruction volume in the x and y directions.

Fig. 4 shows coronal and axial views of images using the proposed artifact subtraction approach in (12). The second and third column show reconstructed images using the SAA and Maximum Likelihood (ML) algorithms. However, since the SBDX system has a very small tomographic angle, neither algorithm provided effective suppression of tomosynthesis artifacts. In the coronal view, the nodule (in the middle) is visible in both reconstructions, but the boundaries are not very clear. The fourth and fifth columns are subtraction images with $k = 2$ and $k = 8$. In the coronal view, the subtraction images show the pulmonary nodules with clearer boundaries than either of the reconstructed images. The subtraction with smaller suppression parameter provides improved nodule visualization, and other lung structure details also appear in the image.

IV. Conclusion

Here we described a method for tomosynthesis artifact suppression using simulated images and artifacts using a BAA analytical model. We proposed a computationally practical algorithm to simulate images and out-of-plane artifacts from a patient specific prior CT image. The accuracy of the analytical model, as an approximation of the SAA reconstructed image, was good when the displacement was within the $\pm$ 10 mm region in the x and y directions. Nodule visibility was significantly improved by subtracting simulated artifacts from the reconstructions, especially in the z direction. The future work of this project is finding fast 3D registration methods to align the images before subtraction.

Acknowledgment

This work is supported by Grant NIH R21 HL098683. The authors wish to thank Geoffery Nelson, Tobias Funk and Sungwon Yoon for their helpful assistance and discussion. The authors are grateful to Dr. Sandy Napel for patient CT data.

References

[1] J. T. r. Dobbins and D. J. Godfrey, "Digital x-ray tomosynthesis: current state of the art and clinical potential.," *Phys Med Biol*, vol. 48, pp. R65–106, Oct 2003.

[2] E. G. Solomon, B. P. Wilfley, M. S. V. Lysel, A. W. Joseph, and J. A. Heanue, "Scanning-beam digital x-ray (sbdx) system for cardiac angiography," vol. 3659, pp. 246–257, SPIE, 1999.

[3] M. A. Speidel, B. P. Wilfley, J. M. Star-Lack, J. A. Heanue, and M. S. Van Lysel, "Scanning-beam digital x-ray (sbdx) technology for interventional and diagnostic cardiac angiography.," *Med Phys*, vol. 33, pp. 2714–2727, Aug 2006.

[4] B. Li, G. B. Avinash, J. W. Eberhard, and B. E. H. Claus, "Optimization of slice sensitivity profile for radiographic tomosynthesis.," *Med Phys*, vol. 34, pp. 2907–2916, Jul 2007.

[5] Y. Schwarz, J. Greif, H. D. Becker, A. Ernst, and A. Mehta, "Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid ct images: the first human study.," *Chest*, vol. 129, pp. 988–994, Apr 2006.

[6] S. Yoon, A. Pineda, E. Solomon, J. Star-Lack, and R. Fahrig, "A fast and accurate tomosynthesis simulation model," in *Nuclear Science Symposium Conference Record, 2004 IEEE*, vol. 6 of *10.1109/NSSMIC.2004.1466746*, pp. 3966 – 3970 Vol. 6, 2004.

Improved CBCT image reconstruction for IGRT

Xiao Han*, Erik Pearson†, Charles A. Pelizzari†, and Xiaochuan Pan*†
*Department of Radiology, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637,
†Department of Radiation and Cellular Oncology, The University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637

***Abstract*—Linear-accelerator (LINAC) integrated, cone-beam computed tomography (CBCT) systems are routinely used for image-guided radiation therapy (IGRT). In recent years, remarkable development has been achieved on optimization-based image-reconstruction algorithms, which have demonstrated advantages over conventional, analytic-based algorithms in a number of scenarios. However, CBCT image quality has yet been explored for its full potential under conditions of practical IGRT interest by use of optimization-based reconstruction. In this work, we investigate and exploit the image-quality potential of CBCT under current IGRT imaging conditions via adapting an advanced, optimization-based algorithm for image reconstruction from physical phantom and patient data collected with imaging protocols of high clinical interest. We carry out characterization studies for qualitatively and quantitatively assessing the merit of the optimization-based algorithm under specific imaging tasks. The results show that appropriately designed optimization-based algorithms can yield CBCT images of improved quality than that of conventional algorithms. The results have the implications of improving the CBCT utility for current clinical IGRT applications, as well as potentially enabling non-standard, novel IGRT applications.**

I. Introduction

Linear-accelerator (LINAC) integrated, cone-beam computed tomography (CBCT) systems [1] are routinely used for image-guided radiation therapy (IGRT), where the treatment beam is guided, monitored, and verified with the aid of CBCT images, for treating cancers in head and neck, thorax, pelvis, and so forth [2]. Although current CBCT image quality can be adequate for some of the IGRT applications, the full image-quality potential remains to be explored. Despite the rapid development of CBCT systems and improved performance of novel hardware components, most clinical CBCT systems still employs the conventional, analytic-based algorithms such as FDK and its variants, for image reconstruction. Recent remarkable development on optimization-based image-reconstruction algorithms has demonstrated potential in numerous imaging configurations, such as sparse-view or low-dose imaging [3, 4]. However, CBCT image-quality has yet been fully explored for its potential under conditions of practical IGRT interest by use of optimization-based reconstruction.

In this work, we investigate and exploit the image-quality potential of CBCT under current IGRT imaging conditions via adapting one of the advanced, optimization-based algorithms for image reconstruction. Specifically, we acquire raw projection data of calibration and anthropomorphic phantoms by using protocols available on a clinical CBCT system. The data are then corrected, by use of software utilities provided by the manufacturer, in the same way that clinical CBCT data are corrected. Then, we apply the adaptive-steepest-descent-projection-onto-convex-sets (ASD-POCS) algorithm, to reconstruct images from the corrected data. Finally, we characterize the reconstruction quality with qualitative and quantitative metrics under the context of specific imaging tasks.

II. Data acquisition

A. Phantoms and CBCT system

We scanned a standard phantom, the Catphan phantom, at three sections for characterizing different image-quality properties. We also scanned a anthropomorphic phantom, the Rando head phantom, and a prostate patient. The CBCT system for data acquisition was an on-board imaging (OBI) system on a Trilogy linear accelerator (Varian Medical Systems). The OBI system consists of an X-ray source and a flat-panel detector, which is mounted on the accelerator gantry orthogonal to the treatment beam. The flat-panel detector has an effective 1024×768 square pixel array, with a pixel size of 388 μm. The distance between X-ray source and detector is 150.0 cm and the X-ray source to isocenter distance is 100.0 cm. The detector, supported by a robotic arm, can be aligned with the source-isocenter axis axis to form a full-fan geometry, or it can be shifted laterally to form a half-fan geometry for enlarging scanning field-of-view (FOV) to accommodate larger body cross-sections such as the pelvis region. We investigate in this work imaging protocols employing both full- and half-fan geometries. Due to the limited space, we show only phantom results of full-fan geometry, and will present additional phantom and patient results of half-fan geometry at the conference.

B. Imaging protocols

We selected two imaging protocols available in the OBI application software [5], which are high-quality head (HQH) and low-dose head (LDH), for acquiring CBCT data. Both protocols employ full-fan geometry, and projection data were collected with 100 kVp X-rays at 360 views over a half-scan range of 200 degrees. The two protocols differ by the tube current and exposure time, with HQH employing 80 mA and 25 ms, and LDH 10mA and 20 ms, resulting in a 10-fold difference in total mAs. For each of the two clinical protocols, two additional, customized protocols were also considered, including a full-scan protocol by expanding the projection data set to include 640 views over 360 degrees, and a sparse-scan protocol of 180 views over 200-degree range by removing every other projection-view. We refer to these additional protocols as HQH-f, HQH-s, LDH-f, and LDH-s, where "f" denotes full-scan, and "s" sparse-scan. Thus, for each phantom we obtained data from six protocols. We also investigated additional imaging protocols, such as the low-dose thorax protocol, and will present the results at the conference.

C. Data preprocessing

We processed the acquired data sets by running them through the processing chain corresponding to each clinical protocol available in the iTools software utility provided by Varian Medical Systems, which includes a sequence of operations for flood-image normalization and corrections for physical factors such as scatter and beam-hardening effect. In this way, the condition of data prepared can be regarded identical to that used by clinical reconstruction, and image-reconstruction algorithm is isolated as the only variable of the study.

III. Image reconstruction

A. Optimization-based reconstruction and parameters

Imaging model A discrete-to-discrete imaging model is used for summarizing the CBCT imaging process:

$$\mathbf{g} = \mathcal{H}\mathbf{f}, \tag{1}$$

where vectors $\mathbf{g}$ and $\mathbf{f}$ of sizes M and N denote the discrete data and the discrete image, and the M-by-N matrix $\mathcal{H}$ models the cone-beam X-ray transform. Complete specification of the imaging model requires a number of parameters, such as the shape and size of the image voxels, which we determine in this work according to current clinical CBCT practice. The axial length of the voxels (i.e., slice thickness) is selected to be 2.5 mm, the default values for all protocols under investigation. On the transverse plane the array is represented by 512 $\times$ 512 voxels covering an FOV of 250 mm diameters. Therefore, the resulting voxels are of cuboid shape with dimension of 0.488 $\times$ 0.488 $\times$ 2.5 mm^3.

Optimization-based reconstruction program To solve the imaging model in Eq. (1), we formulate an optimization-based optimization program,

$$D(\mathbf{f}) \leq \epsilon, \;\; ||\mathbf{f}||_{TV} \leq t_0, \;\; f_j \geq 0, \;\; \text{and} \;\; c_\alpha(\mathbf{f}) \leq \gamma, \tag{2}$$

where $D(\mathbf{f}) = |\mathcal{H}\mathbf{f} - \mathbf{g}|/M$ denotes the average Euclidean data divergence per detector pixel between measured data and imaging model; $||\mathbf{f}||_{TV}$ the image's total variation (TV) (see, e.g., Eq. (9) of Ref. [6]); parameter t_0 constrains TV of the reconstructed image; f_j indicates the value at voxel j of image $\mathbf{f}$, $j = 1, 2, ..., N$; and parameter γ constrains the metric $c_\alpha(\mathbf{f})$ (defined in Eq. (21) of Sidky and Pan (2008)). Parameter $\epsilon > 0$ is designed to account for the inconsistency between the measured data and the imaging model.

ASD-POCS algorithm We employ the ASD-POCS algorithm to reach the solution set designed by the reconstruction program in Eq. (2). The framework of the ASD-POCS algorithm, and numerical techniques for monitoring algorithm convergence have been described in detail elsewhere [6], and we focus in the current work on adapting the algorithm to reconstructing CBCT images from data of clinically relevant conditions.

B. FDK algorithm and parameters

We also reconstructed images from each of the prepared data sets by using the FDK algorithm, because FDK is currently the algorithm of choice in most clinical CBCT scanners. We use the FDK reconstruction as the benchmark for the image quality typically seen in clinical CBCT. For this purpose, we selected the Hann filter as the reconstruction kernel for our FDK implementation. It is likely that the image quality of FDK can improve for some of the studied tasks by carefully adjusting reconstruction kernel parameters. However, such an effort is beyond the scope of the work, and we chose the Hann filter because it yields images of appearance close to that of typical clinical reconstructions.

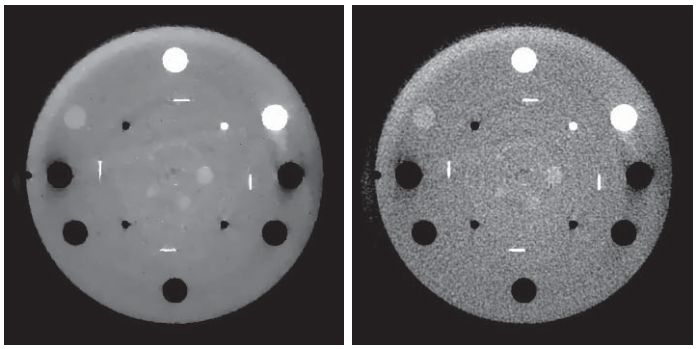


Figure 1. Images of the Catphan phantom within the CTP404 section reconstructed by use of the ASD-POCS (left) and FDK (right) algorithms from data acquired with the HQH protocol.

IV. Results

A. Catphan CTP404 result

The section of CTP 404 contains both high- and low-contrast structures, which resembles a realistic scenario in IGRT where both boney structures and soft tissues (e.g., tumor) are present and used for positioning adjustment and tumor localization.

We reconstructed images using the ASD-POCS algorithm from data acquired with the three HQH-based and three LDH-based protocols. We display in Fig. 1 the image reconstructed from data acquired with the HQH protocol within the ROI including only the central region of the phantom. As a reference we display the FDK reconstruction with Hann kernel. It can be observed that the ASD-POCS algorithm yields a reconstruction of moderately sharper ramps and substantially improved contrast-noise-ratio (CNR)/detectability than the FDK counterpart. Additional images and quantitative characterization results will be presented at the conference.

B. Catphan CTP515 result

The section of CTP515 contains mainly low-contrast inserts, which resembles a realistic scenario in IGRT where visibility of soft tissues is the priority of imaging, whereas moderate compromise on spatial resolution may be acceptable.

We reconstructed images using the ASD-POCS algorithm from data acquired with the three HQH-based and three LDH-based protocols. We display in Fig. 2 the image reconstructed from data acquired with the HQH protocol within the ROI including only the central region of the phantom. As a reference we display the FDK reconstruction with Hann kernel. It can be observed that

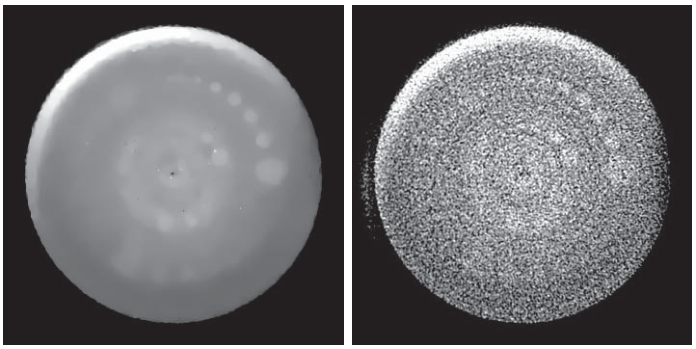


Figure 2. Images of the Catphan phantom within the CTP515 section reconstructed by use of the ASD-POCS (left) and FDK (right) algorithms from data acquired with the HQH protocol.

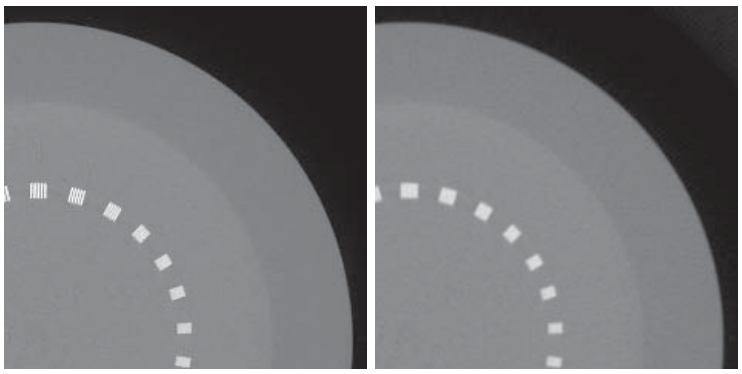


Figure 3. Images of the Catphan phantom within the CTP528 section reconstructed by use of the ASD-POCS (left) and FDK (right) algorithms from data acquired with the HQH protocol.

the ASD-POCS algorithm yields a reconstruction of substantially improved CNR/detectability than the FDK counterpart. Additional images and quantitative characterization results will be presented at the conference.

C. Catphan CTP528 result

The section of CTP528 contains mainly high-contrast bar phantoms, which resembles a realistic IGRT scenario where accurate localization of high-contrast structures is the imaging priority, whereas moderate compromise on noise property may be acceptable.

We reconstructed images using the ASD-POCS algorithm from data acquired with the three HQH-based and three LDH-based protocols. We display in Fig. 3 the image reconstructed from data acquired with the HQH protocol within the ROI including only the top right quarter of the phantom. As a reference we display the FDK reconstruction with Hann kernel. It can be observed that the ASD-POCS algorithm yields a reconstruction of substantially improved spatial resolution than the FDK counterpart. Additional images and quantitative characterization results will be presented at the conference.

D. Additional phantom and patient results

We have reconstructed images from data acquired with Rando head phantom and a prostate patient by using the ASD-POCS and FDK algorithms. The head phantom and patient data help demonstrate the algorithm performance for images of realistic human anatomy. While the head phantom images contain mainly complex boney structures, the prostate patient images contain moderately complex boney structures and abundant soft tissues of low contrast. The results show that ASD-POCS algorithms can yield images of reduced artifacts and improved spatial- and contrast-resolution, which may enhance CBCT's utility for IGRT applications. Additional images and characterization results will be presented at the conference.

V. Discussion

We have investigated the potential room for improvement of CBCT image quality by optimization-based reconstruction under conditions of clinical IGRT interest. Based upon a discrete-to-discrete imaging model, the appropriately designed optimization-based reconstruction has shown the flexibility of incorporating effective data and image constraints, which can help to yield reconstructions of improved quality for imaging tasks under consideration. The ASD-POCS algorithm considered in the work represents one of such optimization-based algorithms, albeit it is likely that other algorithms, when appropriately designed and implemented, can also yield images of comparable or improved quality. Therefore, our purpose is not promoting any particular algorithm. Rather, we have demonstrated the existence of potential in CBCT image quality that can be exploited by advanced optimization-based reconstruction algorithms. The results have implications of improving the CBCT efficacy for current clinical IGRT applications as well as potentially enabling non-standard, novel IGRT applications.

References

[1] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 53, pp. 1337–1349, 2002.

[2] L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T. Li, G. Kim, G. Luxton, A. Koong *et al.*, "Overview of image-guided radiation therapy," *Medical Dosimetry*, vol. 31, no. 2, pp. 91–112, 2006.

[3] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Phys. Med. Biol.*, vol. 55, pp. 6575–6599, 2010.

[4] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, E. L. Ritman, and X. Pan, "Algorithm-enabled low-dose micro-CT imaging," *IEEE Trans. Med. Imaging*, vol. 30, pp. 606–620, 2011.

[5] S. Kim, S. Yoo, F. Yin, E. Samei, and T. Yoshizumi, "Kilovoltage cone-beam CT: Comparative dose and image quality evaluations in partial and full-angle scan protocols," *Med. Phys.*, vol. 37, p. 3648, 2010.

[6] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

Simulation Study of the OpenPET Scanner with Bridge Detectors to Compensate for Incomplete Data

Hideaki Tashima, Taiga Yamaya, *Member, IEEE*, and Paul E. Kinahan, *Fellow, IEEE*

***Abstract*— We are developing an open-type PET "OpenPET" geometry. One possible geometry is a dual-ring OpenPET, which consists of two detector rings separated by a gap for entrance of a radiotherapy beam. In our previous simulations and experiments the OpenPET imaging geometry was shown to be feasible by applying iterative reconstruction methods. However, the gap violates Orlov's completeness condition for accurate tomographic reconstruction. In this study, we propose a solution for the incompleteness problem by adding bridge detectors to fill in parts of the gaps of the OpenPET geometry; we call this bridged OpenPET. Although this geometry was considered previously, its analytical property was not discussed. Therefore, we applied the direct Fourier method as an analytical reconstruction method to the bridged OpenPET, dual-ring OpenPET and conventional cylindrical PET for comparison. Numerical simulations showed that the additional bridge detectors compensate for the incompleteness of the OpenPET by covering one direction perpendicular to the transaxial slices of the imaging subjects.**

***Index Terms*—Positron emission tomography, OpenPET, Whole-body PET, Image reconstruction**

I. Introduction

We are developing an open-geometry PET scanner, OpenPET, which has axially separated detector rings and a physically opened field of view [1]-[7] (Fig. 1). The OpenPET geometry can reduce patient stress due to claustrophobia during PET brain imaging. The OpenPET scanner also enables various applications such as in-beam PET imaging for particle therapy [8]-[14] and entire body PET imaging using fewer detector rings [4]-[6]. The OpenPET geometry that consists of two detector rings separated by a gap is called the dual-ring OpenPET.

The open space between the detector rings is imaged only from oblique lines of response (LORs), in which low frequency components are lost [15]. There is no LOR that forms direct planes. Thus, the OpenPET image reconstruction is an incomplete inverse problem because Orlov's condition is not satisfied [16]. However, our previous simulations and experiments showed that it is feasible to obtain the reconstruction images even for the in-gap region by using iterative image reconstruction methods such as the maximum likelihood expectation maximization (MLEM) which is the most commonly used iterative method for PET [1]-[5]. The effect to compensate for the incomplete data in the iterative methods was proven by applying the method of convex projections to analytically reconstructed OpenPET images [17]. In this study, we propose an alternative approach for the incomplete problem by adding bridge detectors to fill in parts of the gaps of the OpenPET geometry (Fig. 1). We call the OpenPET geometry with the bridge detectors connecting detector rings on both sides "bridged OpenPET". To demonstrate the effectiveness of the bridged OpenPET, first we discussed theoretical aspects of the geometry in comparison with the dual-ring OpenPET and conventional cylinder PET. Next, we conducted numerical simulations by using the direct Fourier method [18] as an analytical method to evaluate the imaging performances of these geometries.

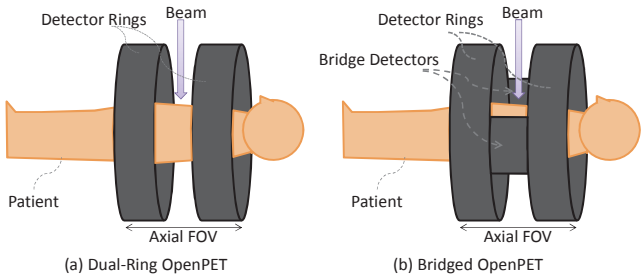


Fig. 1. Schematic illustrations of the dual-ring OpenPET and bridged OpenPET enabling radiation therapy during PET scanning and extension of the axial FOV.

II. Theory

Image reconstruction by the direct Fourier method requires 2D projection data. Each 2D projection set is calculated for parallel LORs. If there are gaps in the 2D projection data, the reconstructed image will have artifacts. To simplify the problem, we used only projection angles without truncation as shown Fig. 2. Then, we focused on the imaging problem of the in-gap region for the OpenPET geometries, where the feasible imaging region is the rhombus region in Fig. 2. For the dual-ring OpenPET, only two projection angles are available. In addition to these angles, projection angles between 0 to θ are

H. Tashima is financially supported by the Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science.

H. Tashima is with the National Institute of Radiological Sciences, Chiba 263-8555, Japan (e-mail: tashima@nirs.go.jp).

T. Yamaya is with the National Institute of Radiological Sciences, Chiba 263-8555, Japan, and also with the Research Center for Frontier Medical Engineering, Chiba University, Chiba 263-8522, Japan (e-mail: taiga@nirs.go.jp).

P. E. Kinahan is with the University of Washington School of Medicine, Department of Radiology, Seattle, WA 98195 USA (e-mail: kinahan@uw.edu).

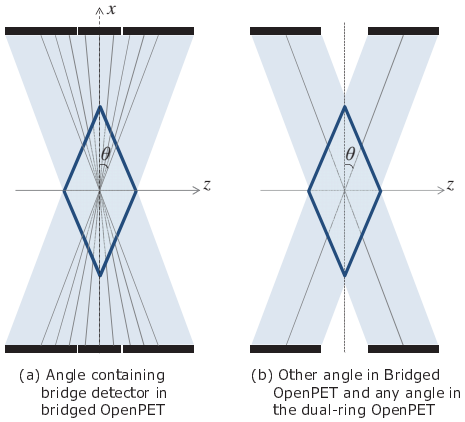


Fig. 2. Available angles without truncation in 2D parallel projections for the rhombus region. The bridged OpenPET can use projection angles between 0 and θ for only one radial angle. For the other radial angle, the available angle is the same as that of the dual-ring OpenPET.

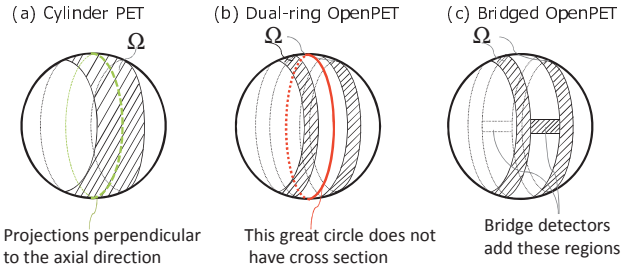


Fig. 3. Illustrations to indicate Orlov's sphere for the center region in each geometry. Orlov's sphere indicates the set of endpoints of normal vectors of the non-truncated 2D parallel projections as Ω.

available for only one radial angle in the case of the bridged OpenPET.

Orlov's condition is useful for determining the data completeness of tomographic applications. We let Ω be the set of endpoints of normal vectors of the non-truncated 2D parallel projections. Orlov's condition states that the image can be reconstructed in a stable way from the set of non-truncated projections if and only if there is no great circle on the unit sphere that does not intersect the set Ω. From the three-dimensional projection-slice (or central-section) theorem this can be interpreted as saying that every portion of the three-dimensional Fourier transform of the object is measured at least once [19]. Since an object and its Fourier transform are one-to-one linear transformations (for square-integrable functions), knowing the entire Fourier transform means the entire object is recoverable. Fig. 3 illustrates three unit spheres showing the set Ω for the center region in conventional cylindrical PET, dual-ring OpenPET, and bridged OpenPET. We note that the set Ω for the dual-ring OpenPET (Fig. 3(b)) is similar to that found in reconstruction problems for the off-center region in the cone-beam X-ray CT [20] and for ectomography [21]-[23]. From these illustrations, we found that the bridged OpenPET satisfies Orlov's completeness condition, while the dual-ring OpenPET requires compensation methods to image accurately in the gap region.

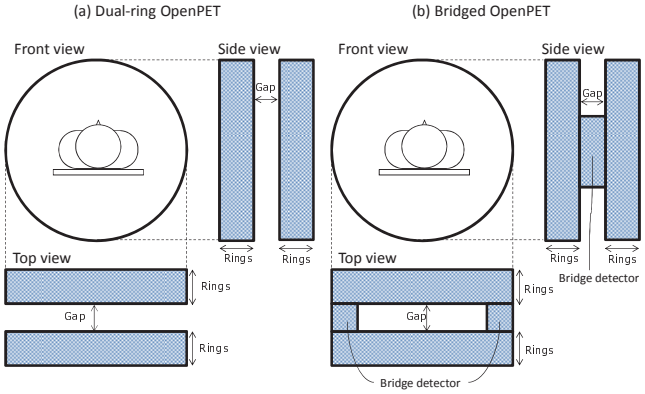


Fig. 4. Simulation geometries for the dual-ring OpenPET (a) and for the bridged OpenPET (b). The bridge detector size is assumed to just cover the imaging object support.

TABLE I
PARAMETERS FOR SIMULATED OPENPET SCANNERS

Parameter	Value
No. of rings	48 (24×2)
Gap	100 mm
Radial sampling	3.125 mm
No. of radial samples	128
No. of angular samples	128
Ring spacing	6.25 mm
Image matrix size	128×128×128
Voxel size	3.125×3.125×3.125 mm^3

III. SIMULATION METHOD

To demonstrate the imaging performance of the bridged OpenPET compared with the dual-ring OpenPET and cylinder PET, we conducted numerical simulations using the geometries shown in Fig. 4. The detailed parameters are given in Table I. Common parameters were used for both geometries and the cylinder PET was assumed to just cover the object support; therefore, only the projection angle parallel to the transaxial plane was used in the direct Fourier image reconstruction method. The bridge detectors were designed to just fit the size of the object support in order to minimize additional cost; therefore, only one radial angle had projection angles between 0 and θ.

We simulated three test phantoms referred to as disk A, disk B, and spots. Each phantom had a cylindrical background region. The diameter of the cylinder was 150 mm and the axial length was 75 mm. The Disk A phantom included three disks with a thickness of 12.5 mm separated by12.5 mm. The Disk B phantom also included three disks but the thickness of each disk was 6.25 mm while the distance between them was 6.25 mm. The Disk A and B phantoms are similar to the Defrise phantom, which is used to evaluate cone-beam artifacts. The spots phantom included 15 spots symmetrically placed as shown in Fig 5.

The projection data were generated by forward projection assuming detector rings and they were resampled into the plane

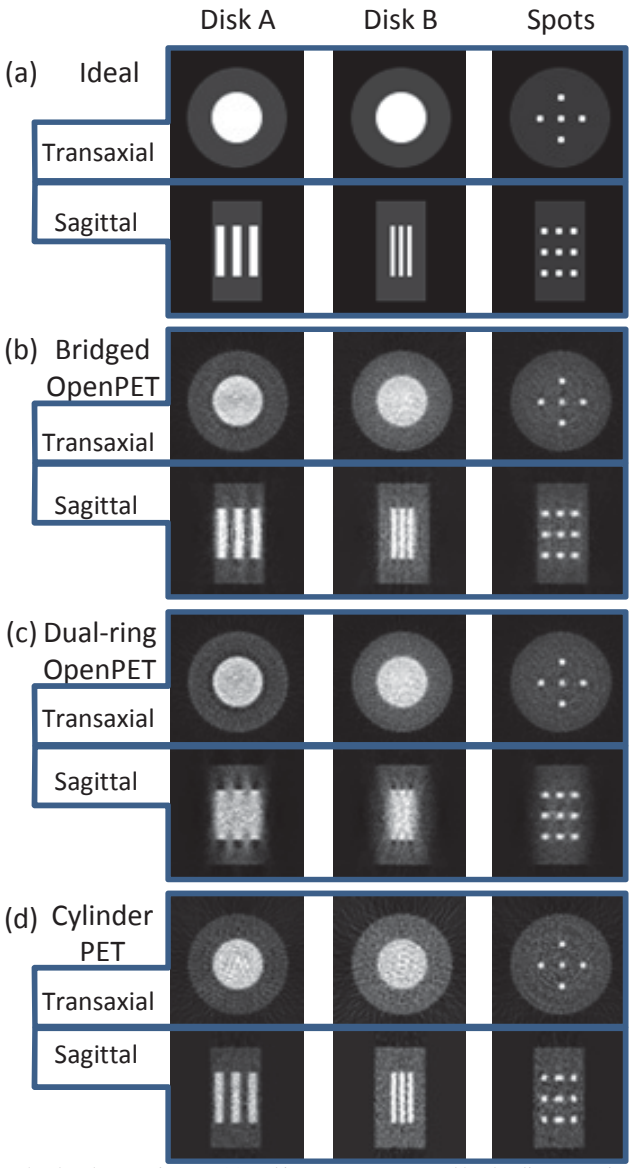


Fig. 5. The phantom images (a) and images reconstructed by the direct Fourier method for the bridged OpenPET (b), the dual-ring OpenPET (c), and the conventional cylindrical PET (d). For each pair, the top is the transaxial slice and the bottom is the sagittal slice. The axial direction is the horizontal direction in the sagittal images.

perpendicular to the projection angle. Poisson noise was added to the projection data and the noise level was around 5 %. The projection data were reconstructed by the direct Fourier method. The values in the truncated region in the Fourier domain for the dual-ring OpenPET was set to zero.

IV. Results

Fig. 5 shows the reconstructed images and Fig. 6 shows the sagittal slices of the Disk B phantom. Fig. 7 shows the center profiles of the reconstructed images of the Disk B and Spots phantoms. The reconstructed images for the dual-ring OpenPET suffered from severe image artifacts along the axial direction. The reconstructed disk phantom images for the bridged OpenPET could significantly reduce the occurrence of image artifacts to the same level as the cylinder PET. The spots

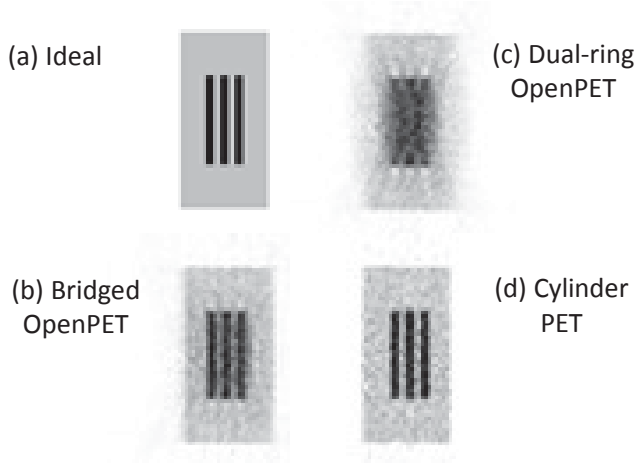


Fig. 6. The sagittal slices of the Disk B phantom in inverse grey scale to show differences and additional features between the geometries.

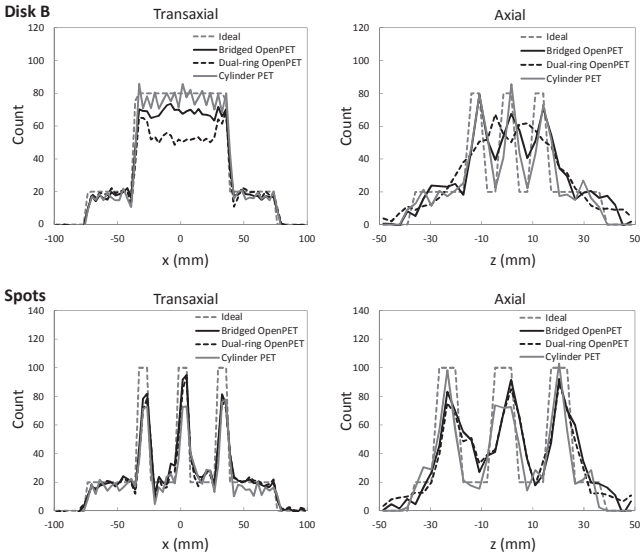


Fig. 7. The center profiles of the reconstructed images of the Disk B (top row) and Spots (bottom row) phantoms.

phantom was clearly reconstructed in all geometries except for the edge of the background cylinder in the image reconstructed for the dual-ring OpenPET.

V. Discussion and Conclusions

The iterative MLEM has been shown to be an effective method for compensating for missing data. However, since MLEM is an non-linear method, its success for arbitrary objects cannot be determined *a priori*. As an alternative strategy to overcome the incomplete data problem in the OpenPET image reconstruction, we evaluated the bridged OpenPET geometry theoretically and numerically. The bridged OpenPET geometry increases the number of detector blocks, resulting in increased costs compared to the dual-ring OpenPET, but the number of blocks is still small compared with the conventional cylindrical PET with the same axial FOV. Furthermore, we showed that the bridged OpenPET geometry could satisfy Orlov's condition. Therefore, we can in principle accurately reconstruct any object in the noise-free case. The numerical simulation showed that

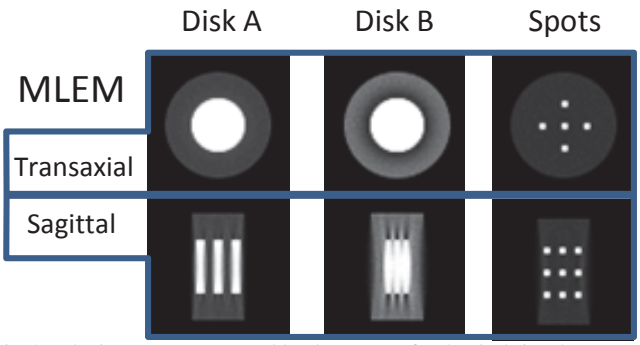


Fig. 8. The images reconstructed by the MLEM for the dual-ring OpenPET shown in the same manner as Fig. 5.

the improvement was substantial for the disk phantoms. Fig. 8 shows the images reconstructed by the MLEM from projection data generated from the same phantoms. The occurrence of the image artifact in the axial direction was suppressed but still existed when the thickness of the disk was thin. It might be possible to further suppress these artifacts if we apply a compressed sensing based approach [24], [25]; however, that is beyond the scope of this article. When we scan subjects containing such structures by the OpenPET, it is worth considering adding bridge detectors to construct the bridged OpenPET. In addition, analytical image reconstruction methods are still preferably used in some areas of study such as brain functional imaging because of their quantitative properties. The bridged OpenPET geometry can be used for such applications if analytical methods or iterative image reconstruction methods in that quantitative performance is ensured are used.

In this study, we assumed that the bridge detector was large enough to cover entire object for avoiding interior problem. Even when the bridge detector is small compared to the object, we expect that the imaging performance will be improved because the data incompleteness is mitigated. Further, there is the possibility of accurate reconstruction for a region of interest (ROI) with *a priori* knowledge for a small region in the ROI [26], [27].

In conclusion, we proposed the bridged OpenPET as an alternative approach which does not have the incomplete data problem. Our theoretical analysis and initial numerical experiment showed that the proposed approach could effectively improve the imaging performance, although at the cost of increasing the number of detector blocks.

REFERENCES

[1] T. Yamaya, T. Inaniwa, S. Minohara, E. Yoshida, N. Inadama, F. Nishikido, et al., "A proposal of an open PET geometry," *Phys. Med. Biol.*, vol. 53, pp. 757-73, Jan. 2008.

[2] T. Yamaya, E. Yoshida, T. Inaniwa, S. Sato, Y. Nakajima, H. Wakizaka, et al., "Development of a small prototype for a proof-of-concept of OpenPET imaging," *Phys. Med. Biol.*, vol. 56, pp. 1123-37, Jan. 2011.

[3] T. Yamaya, T. Inaniwa, S. Mori, T. Furukawa, S. Minohara, E. Yoshida, et al., "Imaging simulations of an 'OpenPET' geometry with shifting detector rings," *Radiol. Phys. Technol.*, vol. 2, pp. 62-9, Dec. 2008.

[4] T. Yamaya, T. Inaniwa, E. Yoshida, F. Nishikido, K. Shibuya, N. Inadama, et al., "Simulation studies of a new 'OpenPET' geometry based on a quad unit of detector rings," *Phys. Med. Biol.*, vol. 54, pp. 1223-33, Mar. 2009.

[5] T. Yamaya, E. Yoshida, N. Inadama, F. Nishikido, K. Shibuya, M. Higuchi, et al., "A Multiplex "OpenPET" Geometry to Extend Axial FOV Without Increasing the Number of Detectors," *IEEE Trans. Nucl. Sci.*, vol. 56, pp. 2644-50, Oct. 2009.

[6] E. Yoshida, T. Yamaya, F. Nishikido, N. Inadama, and H. Murayama, "Basic study of entire whole-body PET scanners based on the OpenPET geometry," *Nucl. Instrum. Methods Phys. Res. A*, vol. 621, pp. 576-580, Apr. 2010.

[7] H. Tashima, E. Yoshida, S. Kinouchi, F. Nishikido, N. Inadama, H. Murayama, et al., "Real-Time Imaging System for the OpenPET," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 1, pp. 40–46, Feb. 2012.

[8] Y. Iseki, H. Mizuno, Y. Futami, T. Tomitani, T. Kanai, M. Kanazawa, et al., "Positron camera for range verification of heavy-ion radiotherapy," *Nucl. Instrum. Methods Phys. Res. A*, vol. 515, pp. 840-49, Dec. 2003.

[9] Y. Iseki, T. Kanai, M. Kanazawa, A. Kitagawa, H. Mizuno, T. Tomitani, et al., "Range verification system using positron emitting beams for heavy-ion radiotherapy," *Phys. Med. Biol.*, vol. 49, pp. 3179-95, July 2004.

[10] T. Nishio, T. Ogino, K. Nomura, and H. Uchida, "Dose-volume delivery guided proton therapy using beam on-line PET system," *Med. Phys.*, vol. 33, pp. 4190-7, Oct. 2006.

[11] W. Enghardt, P. Crespo, F. Fiedler, R. Hinz, K. Parodi, J. Pawelke, et al., "Charged hadron tumour therapy monitoring by means of PET," *Nucl. Instrum. Methods Phys. Res. A*, vol. 525, pp. 284-88, June 2004.

[12] P. Crespo, G. Shakirin, and W. Enghardt, "On the detector arrangement for in-beam PET for hadron therapy monitoring," *Phys. Med. Biol.*, vol. 51, pp. 2143-63, May 2006.

[13] F. Fiedler, G. Shakirin, J. Skowron, H. Braess, P. Crespo, D. Kunath, et al., "On the effectiveness of ion range determination from in-beam PET data," *Phys. Med. Biol.*, vol. 55, pp. 1989-98, Apr. 2010.

[14] T. Inaniwa, T. Kohno, T. Tomitani, E. Urakabe, S. Sato, M. Kanazawa, et al., "Experimental determination of particle range and dose distribution in thick targets through fragmentation reactions of stable heavy ions," *Phys. Med. Biol.*, vol. 51, pp. 4129-46, Aug. 2006.

[15] E. Tanaka and Y. Amo, "A Fourier rebinning algorithm incorporating spectral transfer efficiency for 3D PET," *Phys. Med. Biol.*, vol. 43, pp. 739-46, Apr. 1998.

[16] S. S. Orlov, "Theory of three-dimensional reconstruction. 1. Conditions for a complete set of projections," *Soviet Physics Crystallography*, vol. 20, pp. 312-4, May 1975.

[17] H. Tashima, T. Katsunuma, S. Kinouchi, M. Suga, T. Obi, H. Kudo, et al., "Restoration of the analytically reconstructed OpenPET images by the method of convex projections," in *Fully Three-Dimensional Meeting*, pp. 112–5, 2011.

[18] M. Defrise and P. E. Kinahan, "Data Acquisition and Image Reconstruction in 3D PET," in *The Theory and Practice of 3D PET*, B. Bendriem and D. W. Townsend, Eds. Dortrecht, Netherlands: Kluwer Academic Publishers, 1998, pp. 11-53.

[19] P.E. Kinahan, M. Defrise and R. Clackdoyle, "Analytic Image Reconstruction Methods," in *Emission Tomography*, M. N. Wernick and J. N. Aarsvold Eds. San Diego, CA: Elsevier Academic Press, 2004, pp. 421-442.

[20] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–807, Sep. 2008.

[21] H. E. Knutsson, P. Edholm, G. H. Granlund, and C. U. Petersson, "Ectomography--a new radiographic reconstruction method--I. Theory and error estimates," *IEEE Trans. Biol. Eng.*, vol. 27, pp. 640–8, Nov. 1980.

[22] S. Dale, P. E. Edholm, L. G. Hellström, and S. Larsson, "Ectomography--a tomographic method for gamma camera imaging," *Phys. Med. Biol.*, vol. 30, pp. 1237–49, Nov. 1985.

[23] H. H. Barret, "Limited-angle tomography for the nineties," *J. Nucl. Med.*, vol. 31, pp. 1688-92, Oct. 1990.

[24] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, pp. 660-663, Feb. 2008.

[25] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2791–805, May 2009.

[26] H. Kudo, M. Courdurier, F. Noo and M. Defrise, "Tiny *a priori* knowledge solves the interior problem in computed tomography," *Phys. Med. Biol.*, vol. 53, pp. 2207-31, May 2008.

[27] M. Courdurier, F. Noo, M. Defrise and H. Kudo, "Solving the interior problem of computed tomography using *a priori* knowledge," *Inverse Problems*, vol. 24, p. 65001, Sep. 2008.

Memory access optimization for iterative tomography on many-core architectures

Wim van Aarle, Pieter Ghysels, Jan Sijbers and Wim Vanroose

***Abstract*—Iterative tomographic reconstruction methods, despite their virtues, are known to be slow compared to analytic reconstruction methods, mainly because of the computationally very intensive forward and backward projection operations. By relying on many-core architectures with large vector registers, modern high performance computing (HPC) systems can offer relief. However, to optimally benefit from such systems, the peak performance of the algorithms should not be bound by the memory bandwidth. In this work, a strategy is proposed that improves the performance of the tomographic forward projection by optimizing its memory accesses. Data locality is exploited to hide data access latency and knowledge of the cache architecture is used to optimally distribute the projection operation over many computing cores. Experiments performed on the recently introduced Intel® Xeon Phi™ architecture confirm a substantial boost in projection performance.**

***Index Terms*—Computed Tomography, High Performance Computing, vectorization, many-core computing, Xeon Phi.**

I. Introduction

ADVANCES in tomographic reconstruction techniques continue to lead to significant improvements in reconstruction quality with an ever decreasing radiation dose. Typically however, the price to pay for these improvements is a vastly increased computation time. GPU computing has already been widely applied to alleviate this downside [1], but tomographic algorithms are also ideal algorithms for implementation on general purpose *high performance computing (HPC)* systems. The performance of high-end HPC systems keeps on increasing exponentially; it is expected that by the early 2020s the first machines capable of performing one exaflop (=10^{18} floating point operations) per second will start appearing. To reach that goal, architecture manufacturers can no longer rely on ever increasing clock frequencies — power consumption has become a limiting factor — but are quickly introducing systems with an increasing number of computation cores, each with vector instructions for increasing vector lengths.

Unfortunately, memory performance is not keeping up with processor performance. Furthermore, as more cores are performing computations simultaneously, more data must be transferred from, to, and between these cores. Consequently, data intensive algorithms reach their optimal performance only if they are well adapted to the underlying system architecture.

A good measure to quantify algorithms is their associated *arithmetic intensity*. For each algorithm, this is typically a fixed number and is defined as the number of *floating point operations (flops)* executed per byte fetched from main memory.

Wim van Aarle, Pieter Ghysels and Wim Vanroose are with the Applied Mathematics and Numerical Analysis group at the University of Antwerp, Antwerp, Belgium; and with the Intel ExaScience Lab at IMEC, Leuven, Belgium. Wim van Aarle and Jan Sijbers are affiliated with iMinds-Visionlab, University of Antwerp, Antwerp, Belgium. Contact: wim.vanaarle@ua.ac.be

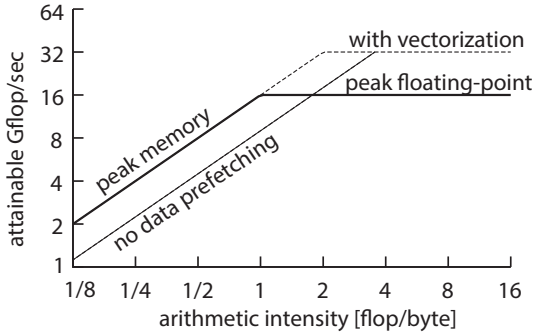


Fig. 1: Roofline model for the maximum attainable floating point performance for numerical algorithms as function of the arithmetic intensity [2]. The numerical values shown here are only an indication. In practice, they depend on the architecture.

The *roofline model*, illustrated in Fig. 1, predicts the maximum attainable performance of a computer algorithm, measured in flops per second, as a function of its arithmetic intensity [2]. The performance of algorithms with a low arithmetic intensity is memory bandwidth bound, while algorithms with a high arithmetic intensity are bound by the performance of the core processing unit.

Tomographic (back-)projection is typically memory bandwidth bound as the resulting code does not perform enough floating point operations per byte fetched from memory to hide the data access latency due to the limited memory bandwidth. The program is then often waiting for data to arrive from main memory. Each data access from main memory can take many processor cycles. With data prefetching, large chunks of data, called *cache lines*, are simultaneously brought closer to the computing processor before the data is even requested. By carefully exploiting data locality in the algorithm, data access latency can thus be hidden.

For algorithms with a sufficiently high arithmetic intensity, the attainable performance is limited by the number of flops each processing core can perform in a given time. As can be seen in Fig. 1, for such cases the peak performance can be substantially increased by vectorizing the program code, e.g. with *single instruction multiple data (simd)* instructions. For example, the recently introduced Intel® Xeon Phi™ architecture supports 512-bit vector registers, allowing one instruction to simultaneously process 16 single precision floating points.

In this paper, high performance computing optimizations are applied to the tomographic forward projection operation. Section II suggests an approach that exploits data locality to increase the benefit of code vectorization. In section III, an approach is suggested to distribute the algorithm on many-core systems in such a way that the bandwidth usage is minimal. Ultimately, Section IV concludes this work.

II. Vectorization of the forward projection

Let $\boldsymbol{v} = (v_j) \in \mathbb{R}^n$ denote a discretized square image of an object, stored in row-major form, with n, the number of pixels. In a 2D parallel beam geometry, projections of $\boldsymbol{v}$ are measured along the lines $x\cos\theta + y\sin\theta = t$, where $\theta \in [-45°, 135°)$ represents the angle between the line and the y-axis and t represents the displacement, or *detector offset*, of the line.

Let m denote the total number of measured detector values for all angles and let $\boldsymbol{p} = (p_i) \in \mathbb{R}^m$ denote the measured projection data. The forward projection can then be modelled as a linear operator $\boldsymbol{W} : \mathbb{R}^n \to \mathbb{R}^m$, that maps the volume $\boldsymbol{v}$ to the projection data $\boldsymbol{p}$, i.e. $\boldsymbol{p} := \boldsymbol{W}\boldsymbol{v}$. In this projection equation, $\boldsymbol{W}$ is an $m \times n$ matrix where w_{ij} represents the contribution of image pixel v_j to detector value p_i.

The projection weights, w_{ij}, can be modelled in various ways. In [3], an overview is given of different methods. It concludes that Joseph's linear interpolation kernel [4] has a sufficiently high accuracy and that it is well-suited for use in high performance computing. It will therefore be used in the remainder of this paper. In Joseph's method, a distinction has to be made between two types of ray. Define *vertical rays* as those for which $\theta \in [-45°, 45°)$ and define *horizontal rays* as those for which $\theta \in [45°, 135°)$.

A. Ray-driven projection

The ray-driven approach is a commonly used forward projection method in which each ray is cast through the volume, thereby summing the contributions of each pixel as the ray passes through. For a vertical ray i, at each row two pixels are hit, i.e. have a non-zero contribution to the ray. Let j denote the index of the left-most pixel. The weights w_{ij} and $w_{i,j+1}$ can then be computed and used to update p_i with the projection of pixels v_j and v_{j+1}. The order of the loops is thus: (1) direction θ; (2) detector offset t; and (3) volume row (for vertical rays) or column (for horizontal rays).

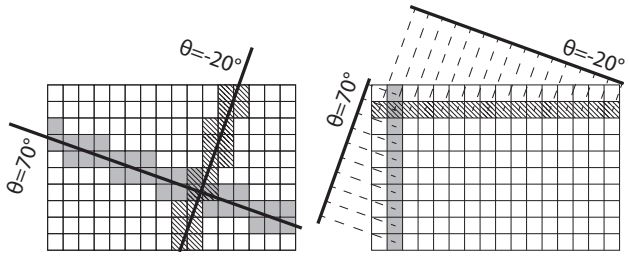
Listing 1: ray-driven projection

```
foreach ray i in [0,m):
   if direction θ of ray i is vertical:
      hitrows ← list of the rows that are hit
      foreach row in hitrows:
         j ← index of left hit pixel
         w_ij ← weight according to Joseph's model
         p_i ← p_i + w_ij v_j + (cos θ − w_ij) v_{j+1}
   else:
      analogue
```

Note that, in Listing 1, only rows are considered that are hit inside the volume window. That way, there are no conditional statements in the inner loop, which would otherwise have prevented its automatic vectorization by modern compilers.

In an optimized C++ implementation, for each byte of data accessed, only about 2 flops are performed. For optimal vectorization performance, it is therefore crucial that the required data reaches the processor as soon as possible, which can be achieved by making good use of data locality. This is accomplished if subsequent data accesses are part of the same cache line, i.e. if they are on the same row. That way, as large chunks of data are prefetched into cache, it is often already available upon request.

In Fig. 4a, the data accesses for a single ray are visualized. It is clear that horizontal rays (e.g. $\theta = 70°$) much more often require data on the same cache line than vertical rays (e.g. $\theta = -20°$). It can therefore be expected that the performance of the projection differs depending on the direction.



(a) ray-driven (b) rayrow-driven

Fig. 4: Visualization of volume data accesses. (a) For a ray-driven approach on row-major data, horizontal rays result in a data access pattern (grey pixels) much more accommodated to data prefetching than vertical rays (striped pixels). (b) For a rayrow-driven approach, vertical rays result in an optimal pattern, whereas horizontal rays result in a worst-case scenario.

B. Rayrow-driven projection

To improve the data access pattern for vertical rays, a reordering of the loops is suggested. In the *rayrow-driven* approach, the loop order becomes: (1) direction θ; (2) row (for vertical direction) or column (for horizontal direction); and (3) detector offset t. A row (vertical direction) or column (horizontal direction) is thus entirely projected in the direction θ before the next row or column is considered.

Listing 2: rayrow-driven projection

```
foreach vertical direction θ:
   foreach row:
      hitrays ← list of the rays that hit the row
      foreach ray i in hitrays:
         j ← index of hit pixel
         w_ij ← weight according to Joseph's model
         p_i ← p_i + w_ij v_j
foreach horizontal direction θ:
   analogue
```

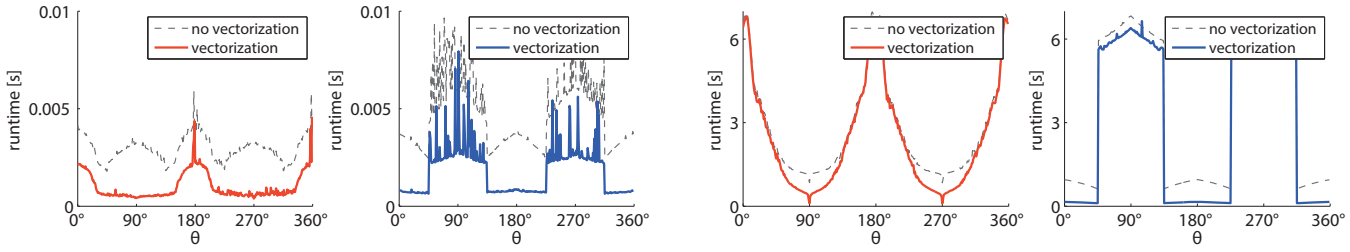
Listing 2 can be implemented such that, for each byte of data accessed only 1 flop is performed.

In Fig. 4b, the data accesses for a single row/column are visualized. For horizontal directions, the data accesses are in fact always optimal and high performance can thus be expected. Vertical directions, however, result in the worst-case scenario in which there is no data locality whatsoever.

C. Hybrid approach

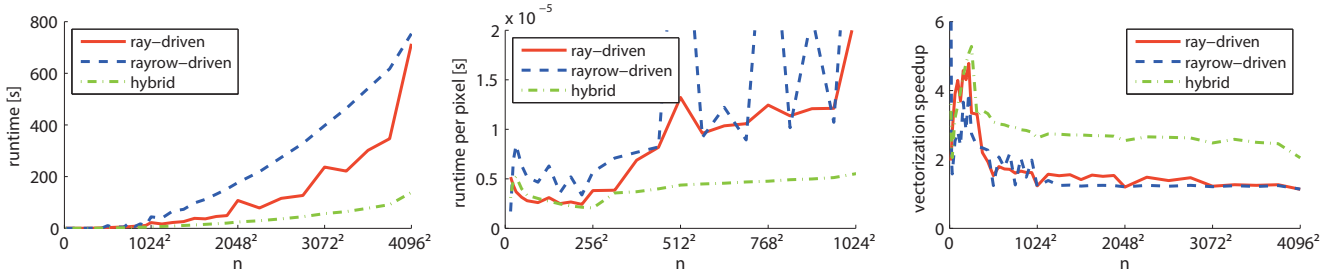
The ray-driven approach shines for horizontal rays, but is far from optimal for vertical rays. For the rayrow-driven approach, exactly the opposite is true. On their own, neither can fully utilize vectorization capabilities of modern architectures.

Fortunately, these two methods are complementary and a *hybrid approach* can easily be constructed by applying the ray-driven approach for horizontal rays, and the rayrow-driven approach for the vertical rays. By selecting the best of both worlds, a great performance improvement can be expected.



(a) ray-driven, $n = 256^2$ (b) rayrow-driven, $n = 256^2$ (c) ray-driven, $n = 4096^2$ (d) rayrow-driven, $n = 4096^2$

Fig. 2: The effect of vectorization for the ray- and rayrow-driven approaches, as a function of the projection angle. (a,b) For volumes that fit into the L2 cache, data accesses are fast enough to benefit from vectorization. (c,d) For larger volumes, vectorization is only useful if data locality can be exploited.



(a) runtime (b) runtime per pixel (c) vectorization speedup

Fig. 3: Experimental results of various projection implementations as a function of the volume size.

D. Experiments

To investigate the performance of the three previously mentioned projection approaches, a series of experiments was performed on a single core of an Intel® Xeon Phi™ "Knight's Corner (KNC)" co-processor. A KNC supports 60 cores, each of which can simultaneously run 4 threads. Combined with 512-bit vector instructions, the peak double precision performance is 1.2 TFlops/s. Each core has 512kb of L2 cache and a stream benchmark measures a bandwidth of 150GB/s. As KNC is an x86-architecture, C++ code can be easily compiled for it. All experiments were performed 5 times of which the median values are presented here.

In Fig. 2, the projection runtime for a single direction is plotted as a function of the angle θ, with vectorization enabled and disabled. Ray- and rayrow-driven approaches are compared for a small volume ($n = 256^2$), and a large volume ($n = 4096^2$). As predicted, both methods are complementary, with the ray-driven method performing well when the rayrow-driven method is slow, and vice versa. For small volumes, vectorization always results in a substantial speedup as the entire volume then fits into the L2 cache and there is little data latency. For larger volumes, a speedup can only be achieved for directions with good data locality.

In Fig. 3a, the projection time of the different methods is plotted as a function of the volume size n. In total, 240 projection directions were used. The hybrid approach clearly offers a substantial performance increase. Fig. 3b shows the projection time divided by the number of pixels in the volume size. It shows that the advantage of the hybrid method is larger for volumes larger than $n = 256^2$, which roughly coincides with the maximal volume size that fits in the L2 cache of a single Xeon Phi computing core. Fig. 3c clearly shows that the hybrid method benefits the most from vectorization.

III. Forward Projection on a Many-Core Architecture

As all rays can be handled independently, forward projection lends itself perfectly to parallel computation on modern many-core architectures. This section investigates the optimal way to distribute the workload over the different cores.

A. Direction-based parallelism

A simple approach to improve performance on many-core systems, is to distribute the outer loop of the algorithms presented in Section II. Each core then computes projection data for one projection direction.

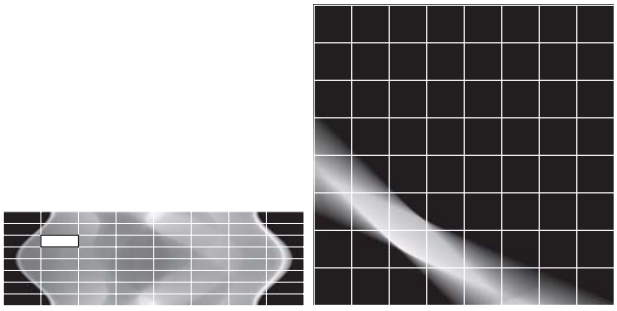
It should be noted that in such an approach, each core needs to access the entire volume data, leading to a vast increase of required memory bandwidth. As such, the conclusions drawn in section II can not be generalized to many-core systems and the performance is not likely to scale well.

B. Patch-based parallelism

In Section II and Fig. 2, it was demonstrated that for volumes that fit entirely into the L2 cache of a core, vectorization is always beneficial. Once the entire volume is loaded into the cache, all data accesses are very fast and do not contribute to the used memory bandwidth. With a *patch-based* projection strategy, this observation is used to improve projection performance, even for very large volumes.

With the patch-based strategy, the projection operation is split into many smaller sub-projections. The projection data is subdivided into a series of patches (Fig. 5a), which can be easily distributed over multiple cores. Furthermore, also the volume data is subdivided into patches, each with a size small enough to fit into the L2 cache (Fig. 5b). That way, the

projection of each small patch can be done very efficiently and substantial performance increases can be expected for large volumes. Also, as the use of memory bandwidth is limited, performance is likely to scale well over multiple cores.



(a) projection data $\boldsymbol{p}$ (b) volume data $\boldsymbol{v}$

Fig. 5: Projection and volume data split into patches. Only a few volume patches contribute to a certain projection patch.

From Fig. 5, it is clear that only a few volume patches actually contribute to a certain projection patch. Therefore, only those patches should be handled, which requires some extra logic at the start of each sub-projection. Also, to increase performance, the data should be stored in *patch-major* form instead of row- or column-major form.

C. Experiments

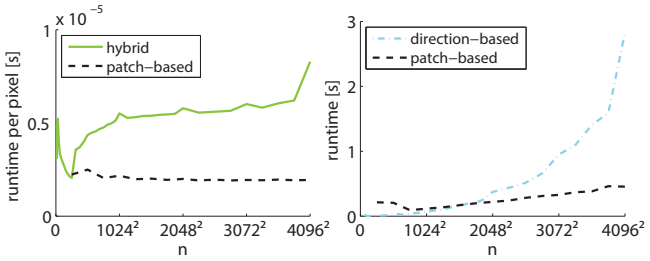
As in section II, a Xeon Phi co-processor is used to demonstrate the patch-based approach. The patch size for the projection data is chosen at 10×256 and that for the volume patches 256×256. The hybrid projection approach is used to project each patch.

Firstly, the hybrid approach is compared with the patch-based projection approach on a single core. Fig. 6a shows the runtime per pixel as a function of the volume size. It should be noted that the hybrid approach has a clear performance decrease as volumes become larger than $n = 256^2$. Many memory accesses are then required as the volume no longer fits into the L2 cache. With a patch-based approach, these memory accesses are limited in number and the runtime per pixel remains constant as the volume size increases.

Secondly, the same experiment is repeated using all 60 cores of the Xeon Phi. The workload is distributed over different cores using the Intel Thread Building Blocks (TBB) library. From Fig. 6b, the substantial performance increase of the patch-based method over the direction-based method, is clear. This can also be seen in Table I, where projection times are listed comparing all projection approaches discussed in this work.

projector type		serial	many-core	speedup
direction-based	ray-driven	714.21s	7.50s	95.2
	rayrow-driven	753.12s	7.57s	99.4
	hybrid	138.99s	2.83s	49.1
patch-based	ray-driven	79.55s	1.41s	56.4
	rayrow-driven	112.60s	2.11s	53.4
	hybrid	32.58s	**0.45s**	72.4

TABLE I: Forward projection times of a 4096×4096 volume with the different approaches discussed in this work.



(a) runtime per pixel, 1 core (b) runtime, 60 cores

Fig. 6: Experimental results of many-core adaptations of the forward projection as a function of the volume size.

IV. Conclusions

In this article, initial efforts were presented that optimize the memory accesses of iterative tomographic methods for use on modern many-core systems.

To reach optimal performance on such systems, vectorization is crucial. It was shown that vectorization is only beneficial for algorithms with a high arithmetic intensity or for algorithms where the data access latency can be hidden by exploiting data locality. For the tomographic projection operation, this can be accomplished by combining a common ray-driven approach, which has good data locality for some directions but bad data locality for others, with a rayrow-driven approach, where the data locality is perfectly complementary to the one of the ray-driven approach.

It was also shown that even with data locality, vectorization can not be achieved if the memory bandwidth is saturated. Therefore, a strategy was proposed to distribute the workload of a forward projection over multiple cores with a limited amount of data transfer. This patch-based method subdivides the projection problem into many sub-problems that are small enough to be stored in the L2 cache of a core.

Experiments performed on an Intel Xeon Phi co-processor, confirm that with the proposed strategies, the tomographic projection operation much more utilizes the capabilities of the architecture, resulting in a substantial performance increase.

Future work will focus on the back-projection, for which the same principles hold, and on algebraic reconstruction techniques.

Acknowledgment

This work is funded by Intel and by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT).

References

[1] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *Journal of structural biology*, vol. 176, pp. 250–253, 2011.

[2] S. Williams, A. Waterman, and D. Patterson, "Roofline: an insightful visual performance model for multicore architectures," *Commun. ACM*, vol. 52, no. 4, pp. 65–76, Apr. 2009. [Online]. Available: http://doi.acm.org/10.1145/1498765.1498785

[3] F. Xu and K. Mueller, "A comparative study of popular interpolation and integration methods for use in computed tomography," in *Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2006*. IEEE, 2006, pp. 1252–1255.

[4] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Trans on Medical Imaging*, vol. 192, pp. 192–196, 1982.

An Algorithm for Interior Tomography

Gangrong Qu

***Abstract*—A algorithm for interior tomography is proposed by constructing a special smooth rectangular window and it can be used to reconstruct interior tomography availably. The complexity of the algorithm is the same as the classical FBP algorithm. The error analysis and the simulations are given.**

I. Introduction

Interior tomography, also called local or region of interest (ROI) image reconstruction, means reconstructing a local region image from the original image projections through the local region plus a fewer extra projections [1]. Interior tomography can reduce the radiation for body and the measurement requirements, also oversize specimens that exceed the scanner field of view could be at least partially imaged [2].

Wavelet-based interior tomographies have been studied [3]–[6], [9].

Recently, significant progresses have been made mainly by Y. Zou and X. Pan, F. Noo and etc., M. Defrise [14], Kudo, G. Wang, H. Yu and Y. Ye et al. in accurate interior tomographies, based on back-projection filter (BPF) method of image reconstruction. Y. Zou and X. Pan established the fan-beam and cone-beam BPF formulae for image reconstruction, based on A.Katsevich's FBP formula of spiral CT [12]. These formulae are accurate for particular local region for interior tomography. Noo et al. also derived the same results independently, using the finite Hilbert transform and FBP formula [13]. Ye, Yu et al. proved that if the image in a subregion of the local region is known, these reconstruction methods have uniqueness theoretically [16], Kudo also established the results independently [15]. Yu et al. gave the simulations [17]. The detail summary for these interior tomographies is in [2].

The TV minimum methods for interior tomography emerge recently [17], [18].

In this article, we propose a FBP algorithm for interior tomography based on constructing a smooth rectangular window, inspired by [8], [9]. The FBP algorithm is simple and easy to implement and its complexity is the same as the classic FBP algorithm.

The paper is organized as follows. In section II, we discuss the FBP algorithm and its convergence. In section II.1, The algorithm for interior tomography is established and the simulations are showed. In section V, we discuss some relative problems.

Gangrong Qu is with the School of Science, Beijing Jiaotong University, Beijing 100044, China. E-mail: grqu@bjtu.edu.cn.

II. FBP Algorithm

Assume that $f(x) \in L^2(R^n)$ is compactly supported. The Radon transform R maps a function $f(x)$ into the set of its integrals over the hyperplanes of R^n. Let S^{n-1} be an unit sphere in R^n. Then for $\omega \in S^{n-1}$ and $p \in R^1$,

$$(Rf)(p,\omega) = \int_{p=x\dot{\omega}} f(x)d\sigma = \int_{\omega^\perp} f(p\omega + y)dy \quad (1)$$

is the integral of f over the hyperplane perpendicular to ω at a distance p from the origin. Rf is an even function on the cylinder $S^{n-1} \times R^1$ of R^{n+1}, i.e., $(Rf)(-p,-\omega) = (Rf)(p,\omega)$ [1].

Theorem II.1. *Assume that the function $F_A(v)$ satisfies*

1) $0 \le F_A(v) \le 1$ *and if* $v \ge \frac{A}{2}$, $F_A(v) = 0$*;*
2) $F_A(v)$ *with respect to* v *is a monotone non-increasing function;*
3) $\lim_{A\to\infty} F_A(v) = 1$.

Let $q_A(u) = 2\int_0^{A/2} v^{n-1} F_A(v)\cos(2\pi vu)dv$. *Then at any continuous point of* $f(x)$

$$\lim_{A\to\infty} \frac{1}{2}\int_{S^{n-1}} d\omega \int_{-\infty}^{\infty} Rf(t,\omega)q_A(x\cdot\omega - t)dt = f(x). \quad (2)$$

Theorem II.1 is proved by the delta sequence of generalized function.

III. FBP Algorithm for Interior Tomography

We construct a smooth rectangular window to implement the interior tomography. $F_A(a)$ is different from the window functions of the classical FBP which is discontinuous at the endpoints of the windows [20]. We will see in the following that the smoothness of $F_A(a)$ is one of the keys for interior tomography. Let

$$h(t) = \begin{cases} e^{\frac{1}{1-|t|^2}}, & |t| < 1 \\ 0, & |t| \ge 1 \end{cases},$$

and $h_\alpha(t) = h(\frac{t}{\alpha})$ with $\alpha > 0$. Then $h_\alpha(t)$ is of C^∞ and its support, supp$h_\alpha \subset [-\alpha, \alpha]$. Assume that $\chi_\alpha(t)$ is the characteristic function of interval $(-1/2+2\alpha, 1/2-2\alpha)$.

$$\chi_\alpha(t) = \begin{cases} 1, & |t| < 1/2 - \alpha \\ 0, & |t| \ge 1/2 - \alpha \end{cases},$$

and $F(t)$ is the constant C multiplying the convolution of $\chi_\alpha(t)$ and $h_\alpha(t)$, $C\chi_\alpha * h_\alpha(t)$.

$$F(t) = C\chi_\alpha * h_\alpha(t) \tag{3}$$

where $C = \int_{-\alpha}^{\alpha} h_\alpha(t)dt$. Then $F(t) = 1$ if $|t| < 1/2 - 2\alpha$, $F(t) = 0$ if $|t| \geq 1/2$ and $F(t)$ is of C^∞. Let

$$F_A(t) = F(\frac{t}{A}). \tag{4}$$

Then $F_A(t)$ satisfies the three conditions of Theorem II.1 and (2) holds.

In two-dimensional case, $\omega = (\cos\theta, \sin\theta)$, and by $(Rf)(-p, -\omega) = (Rf)(p, \omega)$,

$$\int_{|\omega|=1} d\omega \int_{-\infty}^{+\infty} Rf(t,\omega)q_A(x\cdot\omega - t)dt$$

$$= 2\int_0^\pi d\theta \int_{-\infty}^{+\infty} Rf(t,\omega)q_A(x\cdot\omega - t)dt, \tag{5}$$

$$q_A(s) = 2\int_0^{A/2} aF_A(a)\cos(2\pi as)da. \tag{6}$$

By Theorem II.1, at the continuous points of $f(x)$,

$$\lim_{A\longrightarrow\infty} \int_0^\pi d\theta \int_{-\infty}^{+\infty} Rf(t,\omega)q_A(x\cdot\omega - t)dt = f(x). \tag{7}$$

IV. FBP Algorithm for Interior Tomography

Now we discuss the problem of local reconstruction. Suppose $f(x) \in L^\infty(R^2)$ has a compact support and Supp$f \subset \{x||x| < E\}$. Then $Rf(\omega, s) \in L^\infty(Z)$. Let $B(x_0, t) = \{x \mid |x - x_0| \leq t\}$. We reconstruct the image on $B(x_0, R)$ using the projection data $Rf(t, \omega)$ through $B(x_0, R + \tau)$, i.e, $Rf(t, \omega)$ with $|t - x_0 \cdot \omega| \leq R + \tau$ and $\omega = (\cos\theta, \sin\theta)$. Let

$$f_A(x) = \int_0^\pi d\theta \int_{-\infty}^{+\infty} Rf(t,\omega)q_A(x\cdot\omega - t)dt. \tag{8}$$

By (8), we can use $f_A(x)$ as the global reconstructed image. In general, the reconstructed local image has a constant bias besides the error [1]. A method to reduce the bias is to extrapolate the missing projections as the constants [5].

$$(Rf)_{\text{ext}}(t,\omega) = (Rf)(t,\omega) \tag{9}$$

if

$$|x_0|\omega\cdot\omega_0 - (R+\tau) \leq t \leq |x_0|\omega\cdot\omega_0 + R + \tau, \tag{10}$$

$$(Rf)_{\text{ext}}(t,\omega) = (Rf)(R + \tau + x_0\cdot\omega, \omega) \tag{11}$$

if

$$|x_0|\omega\cdot\omega_0 + R + \tau < t < E, \tag{12}$$

$$(Rf)_{\text{ext}}(t,\omega) = (Rf)(-R - \tau + x_0\cdot\omega, \omega), \tag{13}$$

if

$$E < t < |x_0|\omega\cdot\omega_0 - (R+\tau). \tag{14}$$

We use $(Rf)_{\text{ext}}(t, \omega)$ as the projection data to reconstruct the local image. Now for $x \in B(x_0, R)$, the local reconstructed image

$$f_{\text{local}}(x) = \int_0^\pi d\theta \int_{-\infty}^{+\infty} (Rf)_{\text{ext}}(t,\omega)q_A(x\cdot\omega - t)dt$$

$$= \int_0^\pi d\theta \int_{-E}^{E} (Rf)_{\text{ext}}(t,\omega)q_A(x\cdot\omega - t)dt. \tag{15}$$

In (15) the convolution integration is from $-E$ to E because supp $(Rf)_{\text{ext},\omega}(t) = (Rf)_{\text{ext}}(t, \omega) \subset [-E, E]$ by the support theorem of the Radon transform [21]. For $x \in B(x_0, R)$, let $r(x) = f_{\text{local}}(x) - f_A(x)$ be the truncated error function, and $G(t, \omega) = (Rf)_{\text{ext}}(t, \omega) - Rf(t, \omega)$. By (5),

$$r(x) = \frac{1}{2}\int_0^{2\pi} d\theta \int_{|t-x_0\cdot\omega|>R+\tau,|t|<E} G(t,\omega)$$
$$\cdot q_A(x\cdot\omega - t)dt,$$

$$r(x) - r(x_0) = \frac{1}{2}\int_0^{2\pi} d\theta \int_{|t-x_0\cdot\omega|>R+\tau,|t|<E} G(t,\omega)$$
$$\cdot(q_A(x\cdot\omega - t) - q_A(x_0\cdot\omega - t)dt$$

$$= \frac{1}{2}\int_0^{2\pi} d\theta \int_{|t|>R+\tau,|t+x_0\cdot\omega|<E} G(t + x_0\cdot x, \omega)$$
$$\cdot(q_A((x - x_0)\cdot\omega - t) - q_A(-t))dt.$$

Let $q(s) = q_1(s)$. For $s \neq 0$, using integration by parts, we get

$$q(s) = -\frac{2}{(2\pi s)^2} + \frac{C_m(s)}{(2\pi s)^m} \tag{16}$$

where m is a positive integer and

$$C_m(s) = 2\int_0^{1/2} [mF^{(m-1)}(a) + aF^{(m)}(a)]$$
$$\cdot\cos(2\pi as + \frac{m\pi}{2})da. \tag{17}$$

From (6) and (16),

$$q_A(s) = A^2 q(As) = -\frac{2}{(2\pi s)^2} + \frac{C_m(As)}{A^{m-2}(2\pi s)^m}. \tag{18}$$

So, by (18),

$$r(x) - r(x_0) = -\frac{1}{4\pi^2}\int_0^{2\pi} d\omega \int_{|t|>R+\tau,|t+x_0\cdot\omega|<E}$$
$$\cdot G(t + x_0\cdot\omega, \omega)(\frac{1}{(t - (x - x_0)\cdot\omega)^2} - \frac{1}{t^2})dt$$

$$-\frac{1}{A^{m-2}(2\pi)^m}\int_0^{2\pi} d\omega \int_{|t|>R+\tau,|t+x_0\cdot\omega|<E}$$
$$\cdot G(t+x_0\cdot\omega,\omega)(\frac{C_m(A(t-(x-x_0)\cdot\omega))}{(t-(x-x_0)\cdot\omega)^m}+\frac{C_m(At)}{t^m})dt$$
$$=: R_1+R_2. \quad (19)$$

By Shwartz's inequality, $|R_1| \le C_1(R+\tau,R)\|G\|_{L^2(Z)}$, where

$$C_1(R+\tau,R)=\frac{1}{4\pi^2}(\int_0^{2\pi}\int_{|t|>R+\tau}(\frac{1}{(t-R\cos\theta)^2}$$
$$-\frac{1}{t^2})^2 dtd\theta)^{1/2}$$

is a small number [1]. From (17) and Riemann-Lebseque Lemma,

$$\lim_{A\to\infty} C_m(As)=0. \quad (20)$$

If $x \in B(x_0,R)$, then $-R \le (x-x_0)\cdot\omega \le R$, and $\tau - R \le t-(x-x_0)\cdot\omega \le t+R$ for $R+\tau<|t|<E$. So, we assume that

$$|C_m(A(t-(x-x_0)\cdot\omega)|, \quad |C_m(At)| \le \delta_A. \quad (21)$$

Therefore,

$$|R_2| \le \frac{\delta_A\|G\|_{L^\infty(Z)}}{2A^{m-2}(2\pi)^m}\int_0^{2\pi}\int_{|t|>R+\tau,|t+x_0\cdot\omega|<E}$$
$$\cdot(\frac{1}{(t-(x-x_0)\cdot\omega)^m}+\frac{1}{t^m})dtd\theta$$
$$\le \frac{\delta_A\pi\|G\|_{L^\infty(Z)}}{A^{m-2}(2\pi)^m}\int_{|t|\ge\tau}(\frac{1}{t^m}+\frac{1}{(t+R)^m})dt$$
$$\le \frac{4\delta_A\pi\|G\|_{L^\infty(Z)}}{(m-2)A^{m-1}(2\pi)^m}\frac{1}{\tau^{m-1}}.$$

If we use the extra margin of k pixels to reconstruct the image in $x \in B(x_0,R)$, then $\tau = kd$ and $A=\frac{1}{d}$ [20]. So,

$$|R_2| \le 4\inf_{m>2}\frac{\delta_A}{(m-2)k^{m-1}(2\pi)^{m-1}d}\cdot\|G\|_{L^\infty(Z)}.$$

By the analysis above, if we know the value of one point of the original local image, for example, $f(x_0)$, we can eliminate the constant bias, using the following reconstruction formula

$$\bar{f}_A(x)=\int_{|\omega|=1}d\omega\int_{-E}^{E}(Rf)_{\text{ext}}(t,\theta)q_A(x\cdot\omega-t)dt+r(x_0) \quad (22)$$

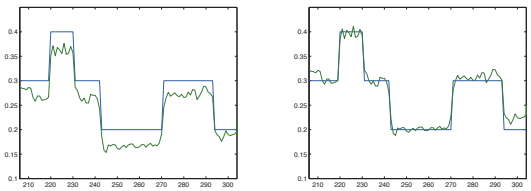
with $r(x_0)=f_{\text{local}}(x_0)-f(x_0)$.

The followings are the stimulation results with $alpha=0.1$.

(a) (b) (c)

Fig. 1

In Fig. 1, (a) is the local original image of radius 50 pixels of 512×512 Shepp-Logan head phantom, (b) is the reconstructed local image by (15) using local data with 2 pixels extra margin and (c) is the reconstructed local image by (22) using local data with 2 pixels extra margin.



(a) (b)

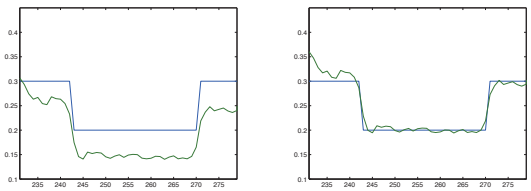
Fig. 2

In Fig. 2, (a) The reconstructed curve picture with pixel $x=256$ compare with the original picture by (15), (b) The reconstructed curve picture with pixel $x=256$ compare with the original picture by (22).

(a) (b) (c)

Fig. 3

In Fig. 3, (a) is the local original image of radius 25 pixels of 512×512 Shepp-Logan head phantom, (b) is the reconstructed local image by (15) using local data with 2 pixels extra margin, (c) is the reconstructed local image by (22) using local data with 2 pixels extra margin.



(a) (b)

Fig. 4

In Fig. 4, (a) The reconstructed curve picture with pixel $x = 256$ compare with the original picture by (15). (b) The reconstructed curve picture with pixel $x = 256$ compare with the original picture by (22)

V. Discussion

The window function $F_A(a)$ in this paper is of C^∞ which is different the window function of FBP in [20]. Therefore the FBP algorithm can be used to reconstruct the interior tomography and its complexity is the same as the classical FBP algorithm. It is easier to implement, comparing with the BPF algorithms. The reconstruction error R_1 is in agreement with the theoretical results of interior tomography [1]. Besides for the error R_1, the unknown constant bias can be reduced by the continuation method in section III and if we know the value of the image but need not know the value of a subregion of the local region, the constant bias can be eliminated.

Acknowledgments

Gangrong Qu is supported partially by national natural science foundations of China (61071144, 61271012).

References

[1] Natterer F 2001 The Mathematics of Computerized Tomography *Classics in Applied Mathematics* (Philadelphia: Society for Industrial and Applied Mathematics)
[2] Clackdoyle R and defrise 2010 M Tomographic reconstrction in the 21th century, *IEEE Signal Magazine* 27 60-80
[3] Berenstein C and Walnut D 1994 Local inversion of Radon transform in even dimensions using wavelets *75 years of Radon transform*, S.Gindikin and P.michor(eds) 45-69 International Press Co. Ltd Cambridge MA
[4] Berenstein C and Walnut D 1996 Wavelets and Local Tomography *Wavelets in Medicine and Biology* Aldroubi, unser (ads) CRC Press
[5] Rashid-Farrokhi F, Liu K J R, Berenstein C and Walnut D 1997 Wavelet-based Multiresolution Local Tomograhpy, *IEEE Trans. On Image Processing* 6 159-166
[6] Zhao S 1999 Wavelet filtering for filtered backprojection in computed tomography *Applied and Computational Harmonic Analysis* 6 346–373
[7] Candès E J 1999 Harmonic analysis of neural networks *Appl. Harmon. Anal.* 6 197-218
[8] Qu G 2000 Inverse Radon Transform with One-dimensional Wavelet Transform *Acta Mathematicae Applicatae Sinica* 16 70-77
[9] Qu G 2006 Wavelet Inversion of the k-plane Transform and its Application *Appl. Comput. Harmon. Anal.* 21 262–267
[10] Walnut D 1992 Application of Gabor and Wavelet Expansions to the Radon Transform *Probabilistic and Stochastic Methods in Analysis, with Application* J.S.Bynes et al (eds) 187-205
[11] Donoho D 2001 Ridge functions and orthonormal ridgelets *J. Approx. Th.* 111 143-179
[12] Zou Y and Pan X 2004 Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT *Phys Med Bio* 49 2004
[13] Noo F, Clackdoyle R and Pack J 2004 A two-step Hilbert transform method for 2D image reconstruction *Phys Med Bio* 49 3903-3923
[14] Defrise M et al 2006 Truncated Hilbert transform and image reconstruction from limited tomographic data *Inverse Problems* 22 1037-1053
[15] Kudo H, Courdurier M, Noo F and Defrise M 2008 Tiny a priori knowledge solves the interior problem in computed tomography *Phys. Med. Biol.* 53 2207C2231
[16] Ye Y, Yu H, Wei Y and Wang G 2007 A General local reconstruction approach based on a truncated Hilbert transform *International Journal of Biomedical Imaging* 63634
[17] Yu H, Yang J, Jiang M and Wang G 2009 Interior SPECT- Exact and stable ROI reconstruction from uniformly attenuated local projections *Communications in Numerical Methods in Engineering* 25 693-710
[18] Yu H and Wang G 2009 Compressed sensing based interior tomography *Physics in Medicine and Biology* 54 2791-2805
[19] Yang J, Yu H, Jiang M and Wang G 2010 High-order total variation minimization for interior tomography *Inverse Problems* 26 035013
[20] Herman G T 1980 Image Reconstruction from Projection New York, Academic Press.
[21] Ludwig D 1966 The Radon transform on Euclidean Space *Com. Pure Appl. Math.* 19 49-81

Reconstruction with Variable Resolution in C-arm Cone-beam CT

Zheng Zhang, Junguo Bian, Xiao Han, Joseph J. Manak, Emil Y. Sidky, and Xiaochuan Pan

Abstract—**In tomographic imaging applications, one is interested often in precise knowledge within a region of interest (ROI) and in rough information outside the ROI. Current optimization-based (i.e., iterative) reconstruction algorithms can yield image values only on uniform grids. It is of practical value to develop iterative algorithms for image reconstruction with variable resolution — a high-resolution ROI image and a coarse image outside the ROI. In this work, we investigate optimization-based algorithms for image reconstruction with variable spatial resolution in C-arm CBCT, and apply the algorithm to data collected with a C-arm CBCT in a head study of patient in which the cerebral ventricle is of interest. The results of our study show that optimization-based algorithms developed can yield an ROI image containing the cerebral ventricles of high spatial resolution, while producing a coarse image outside the ROI. The proposed algorithm can be exploited for lowering computational memory and load and, more important, may reduce artifacts within the ROI resulted from data truncation.**

I. Introduction

A C-arm system with a flat panel detector can be used for collecting CBCT data from which 3D images can be reconstructed. In the last decade or so, a great deal of interest has been developed in using C-arm CBCT as an imaging tool for guiding surgery in operating room (OR) due to its easy accessibility and portability [1], [2]. In contrast to the conventional fluoroscopy mode, CBCT imaging can discern overlapping structures and yield images with improved soft-tissue textures. On the other hand, in comparison to diagnostic CT, C-arm CBCT systems can be of higher flexibility when applied to medical interventional procedures. The unique properties of C-arm CBCT has enabled numerous important imaging techniques for clinical applications, including its imaging of head with low-contrast cerebral ventricles.

C-arm CBCT typically collect data over a short-scan angular range of π plus fan angle, and the FDK algorithm [3] is used to reconstruct images. There has been increased interest in developing optimization-based algorithms for image reconstruction, because they can be more flexible than the FDK and other analytic algorithms in dealing with non-circular geometric and challenging sampling conditions. However, current optimization-based algorithms can reconstruct images on image grids with a single, uniform resolution. This can be time and storage consuming when a fully 3D image is reconstructed from CBCT data. In many applications, however, it is not uncommon that one is interested in fine information only within a region of interest (ROI), while considering rough knowledge outside the ROI acceptable. Therefore, it is of practical value to develop optimization-based algorithms that can yield a reconstruction on grids with variable resolution: fine grids for the ROI, whereas coarse grids for the region outside the ROI.

In this work, we develop a constrained TV-minimization program for the problem, and use the adaptive steepest descend - projection onto convex sets (ASD-POCS) to solve the program [4]–[8]. In particular, we apply the ASD-POCS algorithm developed to reconstructing images from data collected in a patient head scan with a GE C-arm system. For bench-marking the reconstruction, we also use expectation-minimization (EM) and simultaneous algebraic reconstruction technique (SART) to reconstruct images with variable resolution from the data. In the abstract, we include results obtained only with the modified ASD-POCS algorithm and will report results obtained with the other algorithms at the conference.

Z. Zhang, J. Bian, X. Han and E. Y. Sidky are with the Department of Radiology, the University of Chicago.

Joseph J. Manak is with Interventional Systems Laboratory, GE Global Research

X. Pan is with Departments of Radiology & Cellular Oncology, the University of Chicago.

II. Materials and Methods

A. Data acquisition

A C-arm system (GE Healthcare, Buc, France) was used for data acquisition in a brain scan of a patient. The C-arm supports an X-ray source and a flat-panel detector at opposing ends, and rotates them in a circular trajectory. The distance from X-ray source to the rotation axis is 72.4 cm and the distance from X-ray source to the detector is 119.9 cm. The detector is composed of an array of 500×500 elements of size 0.06×0.06 cm^2. The field of view (FOV) is about 18.0 cm within the transverse plane. Projection data of the subject were acquired at 587 views over a short-scan angular range of $193.8°$. Limited by the size of FOV, the data contain transverse truncation from the outer part of the patient head and from the supporting couch. We have performed geometric calibration and artifact correction by using the GE methods [9]–[11]. In particular, the geometric matrix for the actual scanning trajectory was incorporated into the algorithm for image reconstruction.

B. Image arrays of voxels with variable sizes

We reconstruct images using an image array consisting of pixels with different sizes, which is referred to as a variable-resolution array. Such a variable-resolution image array can be designed as follows: the size of voxel within ROI is a and the size in the volume outside ROI is b, as shown on the left panel of Fig. 1. When $a = b$, obviously, the variable-resolution array becomes a conventional, uniform-resolution array with identical voxels. Considering a ROI image with resolution finer than other regions, one can choose $a < b$, and obtain different types of variable-resolution images by controlling the ratio of a to b. Specifically, in this study we fixed $a = 0.04$ cm and considered

different ratios with $b = 0.04$ cm, 0.08 cm, 0.16 cm and 0.32 cm, respectively. The first combination of a and b yields a uniform-resolution image array and others represent different levels of variable resolution. We also investigate impacts of the ratio on ROI image quality.

C. Optimization-based iterative algorithms

The imaging model in CT can be expressed as a discrete linear system [4]–[6]:

$$\mathbf{g}_0 = \mathcal{H}\mathbf{f}, \tag{1}$$

where we use an M-component vector $\mathbf{g}_0$ to represent the model data and an N-component vector $\mathbf{f}$ to represent the image. $\mathcal{H}$ is an $M \times N$ system matrix modeling the X-ray transform.

We formulate the image reconstruction as a constrained optimization problem:

$$\mathbf{f}^* = \operatorname{argmin} \| \mathbf{f} \|_{\mathrm{TV}} \quad \text{s.t.} \quad D(\mathbf{f}) \leq \varepsilon \quad \text{and} \quad \mathbf{f} \geq 0. \tag{2}$$

In Eq. 2, $\| \mathbf{f} \|_{\mathrm{TV}}$ denotes the ℓ_1-norm of the discrete gradient magnitude of the image, and is referred to as the image total-variation (TV).

$$D(\mathbf{f}) = | \mathcal{H}\mathbf{f} - \mathbf{g} | \tag{3}$$

denotes the Euclidean-data divergence, and $\mathbf{g}$ indicates the measured data.

We use the ASD-POCS algorithm to solve the optimization program in Eq. 2. If necessary, we use data distance gradient descent instead of POCS to further approach the necessary condition derived previously [5].

D. ASD-POCS algorithms for image reconstruction with variable resolution

An analytic-based algorithm is based upon a continuous-to-continuous (CC) imaging model. When applied to discrete data, its portion related to data is discretized. However, the algorithm in such a discrete form can still yield a reconstruction value at any point in the image space. Therefore, images with different resolution can readily be reconstructed by use of analytic-based algorithms. An optimization-based algorithm, however, is based upon a discrete-to-discrete imaging model. Current iterative algorithms are designed only for yielding an image on an array of identical voxels, and because image reconstruction at an iteration needs information of image of the previous iteration, the forward- and back-projection are calculated on the array with identical voxels.

The ASD-POCS algorithm is developed originally for uniform image grids. In this work, it is modified to accommodate image grids with voxels of variable sizes. In this case, image array $\mathbf{f}$ can be expressed as combination of two vectors, $\mathbf{f} = \mathbf{f}_H + \mathbf{f}_L$, where $\mathbf{f}_H$ and $\mathbf{f}_L$ denote high- and low-resolution image array, respectively. The ASD-POCS algorithm consists of two major procedures: POCS and TV gradient descent. In POCS, elements of system matrix $\mathcal{H}$ are different for $\mathbf{f}_H$ and $\mathbf{f}_L$, because $\mathcal{H}$ can be considered to represent the forward-projection process and each element is weighted by the corresponding voxel size. Therefore, elements of $\mathcal{H}$ acting on $\mathbf{f}_H$ and $\mathbf{f}_L$ are weighted very differently due to different voxel sizes. Similarly, when calculating $\| \mathbf{f} \|_{\mathrm{TV}}$ for conventional uniform-resolution image array, element value of each voxel is calculated using the same weighting factor — the volume size of the voxel. But in a variable-resolution image array, size of voxels can be different and are calculated differently too.

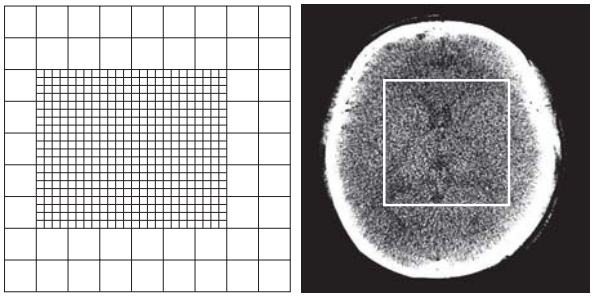


Figure 1. Left: A variable-resolution array. It includes a rectangular-shaped ROI, consisting of pixels of size a, and the region outside the ROI, composing of pixels of size b. Right: Patient head image reconstructed by use of the ASD-POCS algorithm in which ROI is indicated by the white solid lines. The pixel size of 0.04 cm is used in the uniform-resolution array for reconstruction.

III. Results

We have conducted an inverse-crime study to show that, under ideal condition, the proposed algorithm can solve the contrained TV-minimization program formulated for an image array consisting of two regions spanned with voxels of two types of sizes. The results of the inverse-crime study show that the algorithm can solve numerically the program, confirming that the algorithm is a numerically accurate solver of the program. Below, we present results of image reconstruction from real data collected in a patient scan.

We performed ASD-POCS reconstruction on a uniform-resolution array consisting of pixels of size $a = b = 0.04$ cm, which is in the right panel of Fig. 1. The ROI enclosed by white solid lines includes part of the cerebral ventricles, which are related to several brain or mental diseases and thus are a region of highly interest. The ROI in a zoomed-in view is shown on the left panel in the second row of Figs. 2-4, and is used as a gold standard against which ROI reconstructions on variable-resolution arrays are compared. For comparison, we also carried out ASD-POCS reconstructions on uniform-resolution image arrays with three additional pixels of size 0.08 cm, 0.16 cm and 0.32 cm, respectively, and illustrate their ROI images on the right panel in the second row of Figs. 2-4.

We also performed ASD-POCS reconstructions on three different, variable-resolution arrays, illustrated in the middle panel of Figs. 2-4, also with the ROI in zoomed-in view.

Because the contrast of ventricle over its surrounding tissues is quite low, in order to reveal details of the ventricle region, we displayed images at a narrow display window $[0.23, 0.25]$ cm^{-1}. Comparing reconstructions by using the variable-resolution method with the gold standard image in Figs. 2-4, it can be observed that regions outside ROI, with pixel size increasing, lose their resolution rapidly, which causes severe pixelization phenomena. However, there is little impact on the image quality within the ROI, and the texture, contrast and resolution of the structures are comparable to those in ROI of our gold standard. To further demonstrate how the pixel size affects image quality, we compare an ROI image in variable-resolution reconstruction to an ROI image in uniform low-resolution images. On the right

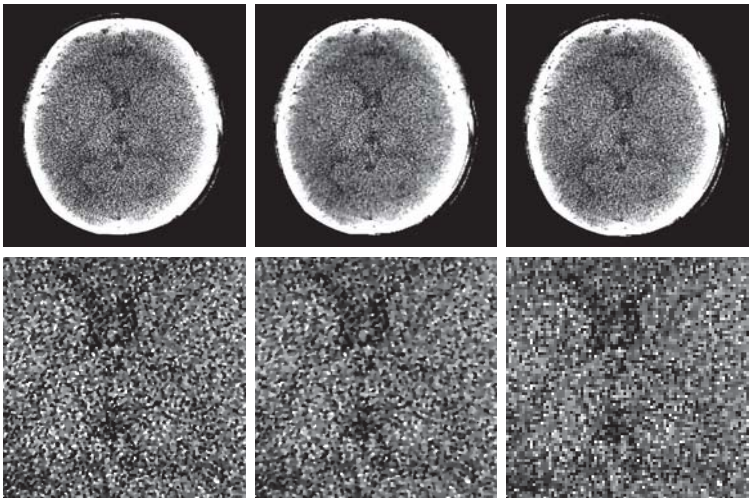


Figure 2. ROI images reconstructed by use of the ASD-POCS algorithm from patient head data on a uniform-resolution array with pixel size 0.04 cm (left), a variable-resolution array specified by the combination of $a = 0.04$ cm and $b = 0.08$ cm (middle), and a uniform-resolution array with pixel size 0.08 cm (right). A display window [0.23, 0.25] cm^{-1} is used.

panel of Figs. 2-4, we observe that as pixel size increases, the ROI image looks more pixelized, the boundaries of the ventricle become blurred and worse-defined, and the contrast of the ventricle deteriorates too. Inspection of the results suggests that the proposed iterative ASD-POCS algorithm can yield a high-resolution ROI that is embedded in a low-resolution image, and image quality within an ROI can be comparable to the ROI of reconstruction on a uniform-resolution array.

IV. Discussions

In this work, we have investigated optimization-based algorithms for image reconstruction with variable resolution. An inverse-crime study was carried out to confirm that the algorithm developed can numerically solve accurately the optimization program formulated for the variable resolution reconstruction. Furthermore, reconstructions from patient data collected with a GE clinical C-arm system indicate that the program and algorithm can yield ROI images with quality comparable to that obtained with uniform, high-resolution arrays. Variable-resolution iterative reconstruction algorithms can be of practical utility because they can reduce computational memory/time and because they may potentially be able to reduce impact of data truncations on ROI images.

References

[1] D. Jaffray, J. Siewerdsen, G. Edmundson, J. Wong, and A. Martinez, "Flat-panel cone-beam CT on a mobile isocentric C-arm for image-guided brachytherapy," *Proc. SPIE*, pp. 209–217, 2002.

[2] J. Siewerdsen, D. Moseley, S. Burch, S. Bisland, A. Bogaards, B. Wilson, and D. Jaffray, "Volume CT with a flat-panel detector on a mobile, isocentric C-arm: Pre-clinical investigation in guidance of minimally invasive suigery," *Med. Phys.*, vol. 32, pp. 241 – 254, 2005.

[3] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, pp. 612–619, 1984.

[4] E. Y. Sidky, K.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. and Technol.*, vol. 14, pp. 119–139, 2006.

[5] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[6] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Probl.*, vol. 25, p. 123009, 2009.

[7] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT ," *Phys. Med. Biol.*, vol. 55, pp. 6575–6599, 2010.

[8] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, E. L. Ritman, and X. Pan, "Algorithm-enabled low-dose micro-CT imaging," *IEEE Trans. Med. Imag.*, vol. 30, pp. 606–620, 2011.

[9] S. Hoppe, F. Noo, F. Dennerlein, G. Lauritsch, and J. Hornegger, "Geometric calibration of the circle-plus-arc trajectory," *Phys. Med. Biol.*, vol. 52, pp. 6943–6960, 2007.

[10] K. Wiesent, K. Barth, N. Navab, P. Durlak, T. Brunner, T. Schuetz, and W. Seissler, "Enhanced 3-D-reconstruction algorithm for c-arm systems suitable for interventional procedures," *IEEE Trans. Med. Imag.*, vol. 19, pp. 391 – 403, 2000.

[11] X. Li, D. Zhang, and B. Liu, "A generic geometric calibration method for tomographic imaging systems with flat-panel detectors," *Med. Phys.*, vol. 37, pp. 3844 – 3854, 2010.

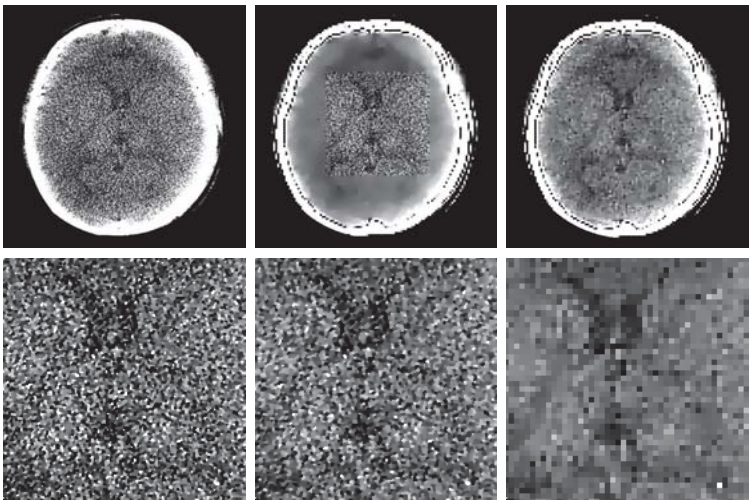


Figure 3. ROI images reconstructed by use of the ASD-POCS algorithm from patient head data on a uniform-resolution array with pixel size 0.04 cm (left), a variable-resolution array specified by the combination of $a = 0.04$ cm and $b = 0.16$ cm (middle), and a uniform-resolution array with pixel size 0.16 cm (right). A display window [0.23, 0.25] cm^{-1} is used.

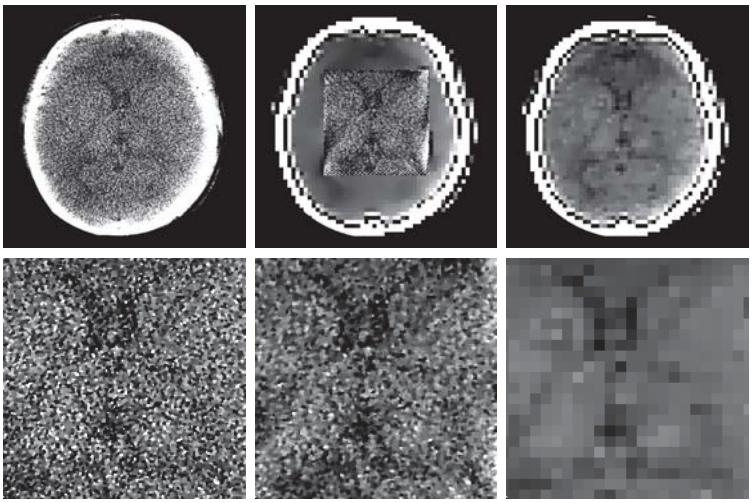


Figure 4. ROI images reconstructed by use of the ASD-POCS algorithm from patient head data on a uniform-resolution array with pixel size 0.04 cm (left), a variable-resolution array specified by the combination of $a = 0.04$ cm and $b = 0.32$ cm (middle), and a uniform-resolution array with pixel size 0.32 cm (right). A display window [0.23, 0.25] cm^{-1} is used.

Blind Deconvolution and SIRT for Cell CT: Contrast Enhancement in Brightfield Microscopic Optical Tomography

Brian Ashcroft, Roger H. Johnson and Deirdre R. Meldrum, *Fellow, IEEE*

Abstract— **Optical tomography of mammalian cells has the potential to open new avenues into understanding subcellular organization and cell structure, function and behavior. We present an implementation of cell-level optical tomography called cell CT. The advantages of cell CT include isotropic spatial resolution, the combination of structural absorption imaging with functional fluorescence imaging and cost effectiveness. The primary limitation on resolution and contrast in the projections acquired from the instrument is the extended point spread function. We propose a simultaneous iterative reconstruction technique (SIRT) to deconvolve each individual projection, using the forward projection to improve the quality of the blur estimate, and then updating the reconstruction volume with the improved true image. This technique allows the cell CT pseudoprojections to be treated as pencil beam projections and also could allow super-resolution optical tomography by the use of multi-frame image reconstruction.**

***Index Terms*— computed tomography, reconstruction algorithms, iterative methods, biomedical optical imaging, optical microscopy, medical diagnostic imaging, biological cells.**

I. Introduction

We propose a robust method to produce tomographic reconstructions of cells by the use of a modified brightfield microscope. The images are improved using a blind deconvolution algorithm that is tolerant of the aberrations inherent to the image formation process in a brightfield microscope. We show that this method results in reconstructions with improved resolution and contrast for small absorbing features in the cellular environment. The technique may also improve the resolution limit of images of absorption-stained samples.

Supported by the W.M. Keck Foundation grant "Live Cell Computed Tomography Instrumentation Development" (D. Meldrum, PI).

B. Ashcroft is with the Center for Biosignatures Discovery Automation and the Center for Single Molecule Biophysics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287 USA (email: brian.ashcroft@asu.edu)

R. H. Johnson is with the Center for Biosignatures Discovery Automation, Biodesign Institute, Arizona State University, Tempe, Arizona 85287 USA (email: roger.h.johnson@asu.edu)

D. R. Meldrum is ASU Senior Scientist; Director, Center for Biosignatures Discovery Automation, Biodesign Institute; and Professor in the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287 USA (email: deirdre.meldrum@asu.edu)

Because of its ability to provide fully 3D images, tomography has revolutionized the study of biology and the diagnosis of diseases[1]. Optical tomography provides an important addition to medical imaging[2]. Cellular optical projection tomography (Cell-CT™, VisionGate, Inc., Phoenix, Arizona) has been developed as a new imaging modality poised to facilitate significant advances in biological imaging and cytological diagnosis[3]. The Cell-CT augments the standard brightfield microscope with the power of the latest advances in microfluidics, electronic cameras and image reconstruction algorithms, provided in a comfortable and familiar environment for the clinical pathologist.

At the same time, Cell-CT provides a unique set of challenges and advantages, some of which relate to the impact of the novel data collection mode on the choice and design of the reconstruction algorithm. This paper explores several tomographic reconstruction algorithms to determine which method minimizes the artifacts unique to this novel tomographic modality and brings the technique firmly within the realm of pencil beam tomography. We develop an effective method to deconvolve the aberrations from the projection formation process in optical projection tomography, to produce higher quality reconstructions. We invoke a minimal amount of simulation, working within the setting of an actual reconstruction.

II. Pseudoprojection Acquisition

The projection acquisition process is shown schematically in Figure 1 [3]. Cells are embedded in a thixotropic carrier gel that is optically matched to the surrounding capillary (horizontal sidewalls shown). The cells are then moved into the field of view of a high numerical aperture (NA) immersion objective lens for data acquisition. The focal plane is scanned through the volume of the cell, while the illumination is held constant. This process is similar to the method used for optically sectioning an object, except that the infinite number of collected planes is integrated onto the camera, and no focused planes are imaged. While these scanned optical projections are mathematically similar to those formed by parallel, straight-line x-ray pencil beams, the formation process and resulting artifacts and aberrations are distinct to this image formation mode. For these reasons, the projection data are referred to as "pseudoprojections". A full electromagnetic analysis of the image formation has been performed by Coe and Seibel [4], forming a strong theoretical

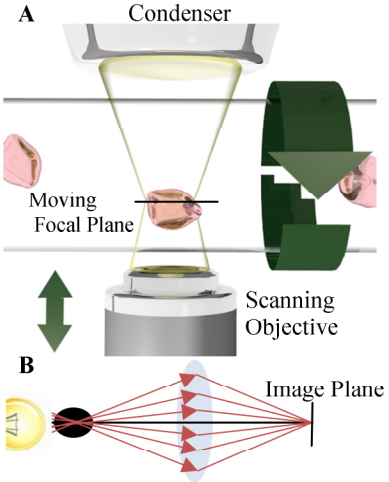


Figure 1. (A) Schematic of the Cell-CT. Cells are fed through a capillary into the field of view of a high NA objective lens. Illuminated by a stationary condenser, the capillary is rotated while the objective is scanned repeatedly through the cellular volume. (B) Ray tracing of cone of rays that interact with the cell at any one optical section. Rays through the focal point add the most weight to the final value at the detector pixel, but a large amount of the cell contributes to this value because the focal plane is scanned.

basis for treating Cell-CT pseudoprojections as approximations of their pencil-beam x-ray counterparts. In order to improve this approximation, we will start with the more intuitive geometric optics to highlight the image formation process and then invoke blind deconvolution to obtain a true pencil beam pseudoprojection.

As shown in Figure 1B, when an optical section is formed, the signal detected at each projection image pixel arises from a cone of rays that pass through the corresponding voxel in object space. The detected optical density with a single ray can be described as

$$d = \exp^{-\int \mu(R)dr} \quad (1)$$

where d is the detected value at the camera pixel and μ is the optical density. The expansion of Equation 1 to a cone that is translated through the object along the Z direction is impractical to integrate; it does show that, for pseudoprojections, almost all the optical density of the imaged object will be added with some weight into each detector pixel.

While geometric optics provides an intuitive description of how the optical image is formed, the formation of an image in an optical absorption microscope is more commonly described as the object density convolved with the 3D point spread function (PSF) of the microscope to avoid the impractical integrals and deal more efficiently with aberrations. This formulation is good enough to produce very high quality images. The image formation is described as

$$v = (V * F) + n \quad (2)$$

where V is the original optical density distribution, v is the measured volume, F is the 3D PSF and n is the noise from the imaging system (normally a mixture of Poisson and Gaussian noise). Assuming an infinite scan range and a finite object (partially false, due to the optical gel), the projection formed in the Cell-CT instrument by sweeping the focal plane through V with the camera shutter open can be described as the 2D projection of this convoluted volume H. This process can be represented as:

$$pp(x,y) = \int_{\mathbb{Z}} (V * F)(x,y,z)dz \quad (3)$$

where pp denotes the 2D pseudoprojection of the convolved volume, $\mathbb{Z}$ indicates integration over z = -∞ to ∞, F is the 3D PSF of the microscope, and V is the true 3D optical density. To reach the pencil beam formulation, Equation 3 is expanded into its integral form:

$$pp(x,y) = \int_{\mathbb{X}}\int_{\mathbb{Y}}\left[\int_{\mathbb{Z}}\int_{\mathbb{Z}} V(x',y',z')F(x-x',y-y',z-z')dz'dz\right]dx'dy' \quad (4)$$

Since V and F are always positive, the integral can be put into a more mathematically convenient form:

$$pp(x,y) = \int_{\mathbb{X}}\int_{\mathbb{Y}}\left[\int_{\mathbb{Z}}|V(x',y',z')|\int_{\mathbb{Z}}|F(x-x',y-y',z-z')|\,dzdz'\right]dx'dy' \quad (5)$$

As the second integral is performed over all possible $\mathbb{Z}$, the offset z' can be ignored, giving

$$pp(x,y) = \int_{\mathbb{X}}\int_{\mathbb{Y}}\left[\int_{\mathbb{Z}}|V(x',y',z')|\,||F||_1 dz'\right]dx'dy' \quad (6)$$

And finally

$$pp(x,y) = \int_{\mathbb{X}}\int_{\mathbb{Y}}||V||_1\,||F||_1 dx'dy' = (PP * f)(x,y) \quad (7)$$

where f is the 2D pencil beam projection of F, and PP is defined as the true pencil beam projection.

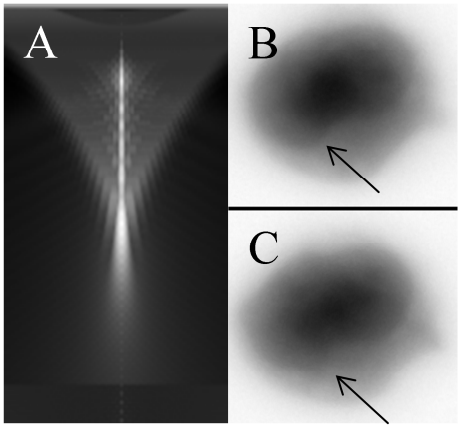


Figure 2. (A) PSF determined by blind deconvolution of an optically sectioned stack acquired by the Cell-CT. The departure from the expected cone is caused by spherical aberrations, which non-intuitively improve tomographic image quality by emphasizing the central values. (B) Comparison of a pseudoprojection with its 180° mirror. Arrows show how the PSF asymmetry in Z changes the features acquired in opposing views.

III. Blind Deconvolution

Many methods have been created to use a known PSF to deconvolve an optically sectioned volume[5, 6]. However, as outlined in Holmes and O'Conner[7], the optical brightfield microscope is prone to a number of effects that make experimental estimation of the PSF problematic [5, 7]. These include phase behavior, scattering, uneven illumination, and

the lack of sub-resolution phantoms that absorb sufficiently to be detected. A further argument for using blind deconvolution within the framework of cell CT can be understood by taking a closer look at pseudoprojection formation. The optical microscope suffers from a number of common aberrations including spherical, astigmatism and coma. As the cell precesses through its rotation, each pseudoprojection is affected by a unique combination of these aberrations.

Blind deconvolution determines the PSF directly from the pseudoprojection, alleviating most of the problems that have been described above. Applying the methods used by Hirsch et al.[8] to Equation 7, with alterations appropriate for brightfield illumination, it is possible to deconvolve the measured pseudoprojection, *pp*, to obtain the true pencil-beam projection, *PP*. Using the pseudoprojection *pp* and its mirror from 180 degrees, pp_m, the true pencil beam pseudoprojection, *PP*, can be determined by using the probability function:

$$p\left(pp, pp_m | PP, f, \tau\right) = \frac{\exp\left\{-\frac{\tau}{2}\left\|(B - pp) - f * PP\right\|^2\right\}}{Z\left(\tau\right)} \quad (8)$$

where $Z(\tau)$ is the partition function, which is dependent on the measurement precision τ[8], and *B* is the brightfield background to help make the deconvolution convex. Estimates of *PP* and *f* are obtained by minimizing the log-likelihood of the cost function

$$\min_{f \geq 0, PP \geq 0} L\left(f, PP\right) = \frac{1}{2}\left\|(B - pp) - f * PP)\right\|^2 \quad (9)$$

As both PP and f will always be positive, this equation can be solved as a nonnegative quadratic programming problem. Sha et al.[9] provide an auxiliary function to solve for *f* and *PP* which leads to an update of the form

$$PP \leftarrow PP \bigodot \frac{\overline{F}^T (B - pp)}{\overline{F}^T \overline{F} PP} \quad (10)$$

for brightfield deconvolution, where $\bar{F}$ is a block-Toeplitz structured matrix; and Θ and the quotient indicate an elementwise multiplication and division. Equation 10 can reversed to find the PSF, *f,* by the commutativity of the convolution. $\bar{F}$ never has to be calculated explicitly as it has the convenient property

$$\overline{F}x = \mathcal{F}^{-1}\left\{\mathcal{F}(f)\mathcal{F}(x))\right\} \quad (11)$$

Without a regularization step, the PSF tends to collapse into a delta function or to spread to unity before reaching the correct solution. Following Holmes [7], a limit was set to the physically allowable values for the PSF. When the PSF wanders from the allowed values, the offending values are reset to the limiting PSF. This allows the PSF to adapt to the irregularities of position and depth without taking on unphysical solutions.

There are a number of effective super-resolution microscope devices for fluorescence imaging, but absorption imaging remains constrained by the diffraction limit. Cell CT might be an attractive platform for achieving enhanced resolution due to the availability of angular oversampling and the fact that many of the bugs of brightfield imaging (such as spherical aberration) become features with the formation of a pseudoprojection. Deconvolution of the pseudoprojections can be utilized to improve the final tomographic reconstruction by performing an image reconstruction while deconvolving. Hirsch devised a variant of this method using blind deconvolution and a resizing matrix to achieve super resolution in an astronomy image series [10]. In this case, the probability function, Equation 8, is altered to

$$p\left(pp, pp_m | PP, f, \tau\right) = \frac{\exp\left\{-1/2\left\|(B - pp) - D_n^{sn}(f * PP)\right\|^2\right\}}{Z} \quad (12)$$

D_n^{sn} is a resizing matrix designed to keep the total power in the image constant:

$$D_n^m = \left(I_n \bigotimes 1_m^T\right)\left(I_m \bigotimes 1_n\right)/n \quad (13)$$

where I_n is the *n x n* identity matrix, 1_n is a column vector of length n and value 1, and ⊗ denotes the Kronecker product.

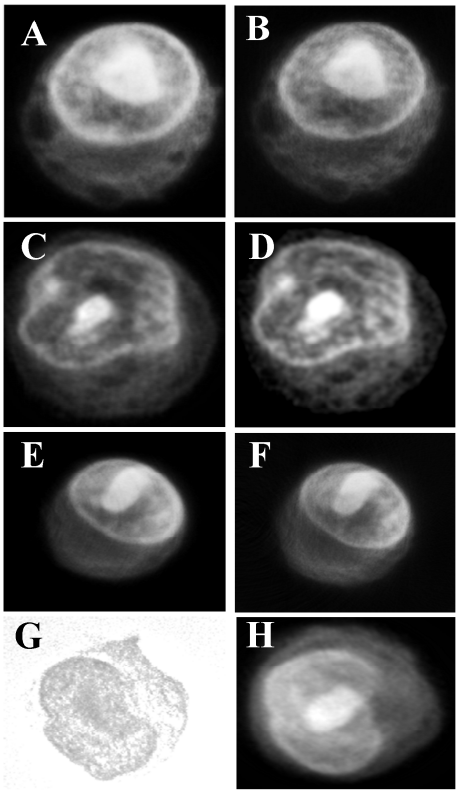


Figure 3. (A) Slice through a FBP reconstruction of an epithelial cancer cell. (B) Same cell reconstructed by the same technique after blind deconvolution of 500 projections. (C) Slice through a SIRT reconstruction of an epithelial cancer cell. (D) The same cell reconstructed by the modified SIRT technique. The positivity constraints result in a modified edge. (E) FBP image used as the phantom for (F) Reconstruction from degraded projections restored to original resolution using super-resolution. (G) Deconvolution of optically sectioned stack image. (H) Tomographic reconstruction of the same cell with blind deconvolution.

IV. Filtered Backprojection Reconstruction

We performed the parallel beam FBP experiments by first subtracting the background, then deconvolving the pseudoprojection with Equation 10. Finally, we reconstructed the 3D volume using a Hann filter and bilinear interpolation. An example result is shown in Figure 3A and B. Fine image

detail is improved by the process, and there is a slight alteration of the noise profile.

We have not yet devised an effective phantom (~10 micron test object containing subresolution absorbers) to validate our results. We instead used a simulated system by reducing the resolution of the input projections to half the original resolution by simple pixel binning. The reconstruction "recovered by deconvolution" was then compared to the original reconstruction to determine whether fine detail was salvaged or lost in the reconstruction from compromised projections.

The multiple frames we used to improve the resolution of the images were the pseudoprojection at one angle and its opposing view. Capillary runout effectively wobbles the camera pixels in both the Z and azimuthal directions, a prerequisite for improving the resolution of the images. As can be seen in Figure 3F, there is the unanticipated result that the restored image seems to have even more fine detail than the original image, which may be a result of the denoising inherent in this method.

V. SIRT Reconstruction

Blind deconvolution of the projections offers a number of advantages for iterative reconstruction methods. Within the framework of SIRT, instead of first completely deconvolving all the pseudoprojections and then backprojecting, the forward and backprojection processes become part of each deconvolution iteration, allowing the determination of successively improved estimates, *PP*, by leveraging information available from the projection set as a whole. The breakdown of the technique is as follows: For each projection angle over 180 degrees, estimate initial *PP* and *f* by using the acquired projection (*pp*) and its mirror twin (180-degree pair). Perform SIRT to create a crude reconstruction using all acquired projections (current estimate of *V*). Forward project *V* to get *PP* and use this image to provide the current estimate for the deconvolution. Refine *f* and *PP* using Equation 10 for several iterations using the *pp* and its mirror twin. The additional benefit of this operation is that *pp* alignment is refined on each iteration. Backproject the difference (error function) between the current forward projection and *pp*. Repeat until a stop condition is reached.

This method is more computationally expensive than standard SIRT methods, but much less intensive than a model reconstruction. The cost must be balanced carefully by pre-estimation of the reconstruction and PSF for each projection. As shown in Figure 3C and D, our SIRT technique produces a reconstruction with improved fine detail. The standard SIRT method provides a higher quality reconstruction than the FBP method due to the high levels of noise rejection. The deconvolution does not reveal additional details, but does provide a better estimate of the chromatin texture, which is important for cancer cell classification.

VI. Conclusion

Blind pseudoprojection deconvolution improves the contrast and resolution of the reconstructions produced by FBP and modified SIRT. The technique is very computationally intensive; however, Figure 3 shows the results to be superior to those obtainable by mere deconvolution of the 3D volume. The method does have a computational advantage over a full model reconstruction due to the total number of rays that must be traced through the volume to deal with the incoherent angular illumination. Care must be taken when interpreting these results as the scheme shows vulnerabilities (instabilities) to some types of artifact that can result from the acquisition process. However, when the acquisition is perfected to the extent possible, the increased conspicuity of fine detail like chromatin texture in the reconstructions is evident.

Acknowledgment

We are thankful to Vivek Nandakumar and Laimonas Kelbauskas for help in operation of the Cell-CT and evaluation of the reconstructions.

References

[1] W. A. Kalender, *Computed Tomography*: John Wiley & Sons, 2011.

[2] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems,* vol. 15, pp. R41-R93, 1999.

[3] M. Fauver, E. Seibel, J. Rahn, M. Meyer, F. Patten, T. Neumann*, et al.*, "Three-dimensional imaging of single isolated cell nuclei using optical projection tomography," *Optics Express,* vol. 13, pp. 4210-4223, 2005.

[4] R. L. Coe and E. J. Seibel, "Computational modeling of optical projection tomographic microscopy using the finite difference time domain method," *Journal of the Optical Society of America A,* vol. 29, pp. 2696-2707, 2012.

[5] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine,* vol. 13, pp. 43-64, 1996.

[6] J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in Astronomy: A Review," *Publications of the Astronomical Society of the Pacific,* vol. 114, pp. 1051-1069, 2002.

[7] T. J. Holmes and N. J. O'Connor, "Blind deconvolution of 3D transmitted light brightfield micrographs," *Journal of Microscopy,* vol. 200, pp. 114-127, 2000.

[8] M. Hirsch, B. Schölkopf, and M. Habeck, "A blind deconvolution approach for improving the resolution of cryo-EM density maps," *Journal of computational biology: a journal of computational molecular cell biology,* vol. 18, pp. 335-346, 2011.

[9] F. Sha, Y. Lin, L. K. Saul, and D. D. Lee, "Multiplicative Updates for Nonnegative Quadratic Programming," *Neural Computation,* vol. 19, pp. 2004-2031, 2007.

[10] M. Hirsch, S. Harmeling, S. Sra, and B. Schölkopf, "Online multi-frame blind deconvolution with super-resolution and saturation correction," *Astronomy & Astrophysics,* vol. 531, pp. A9, 2011.

A Reconstruction Method for Slow Step kVp-Change Multi-Energy CT

Le Shen, Yuxiang Xing, *Member, IEEE*

***Abstract*—The development of dual-energy and multi-energy computed tomography makes it possible to calculate the composition density of different basis materials, or the distributions of electron density and atomic number, which are useful in medical applications and industrial inspections. In this paper, we propose a slow-change kVp multi-energy CT scan scheme (SegMECT) that is convenient to be implemented on a conventional single-energy CT. In our SegMECT, a circular trajectory in a CT scan is angular-equally divided into several arcs for which a certain kVp is chosen. Thus, during the scan, we only need to make a few step changes to the X-ray energy to complete multi-energy data acquisition. In this situation, the image reconstruction problem belongs to the limited-angle problem, which is of practical significance and difficulty. Inspired by the prior image constrained compress sensing (PICCS), we present a prior image based reconstruction techniques to solve this problem. Both numerical simulation and practical experiment are carried out to validate the proposed method and system. The results demonstrate its potential for practical applications.**

***Index Terms*—Iterative reconstruction; Limited angle; Material Decomposition; Multi-Energy; Prior Image;**

I. Introduction

DUAL energy CT (DECT) has been widely utilized in clinical applications due to its ability of material separation. Physically, the linear x-ray attenuation coefficient of a material can be contributed to two major photon interactions: photoelectric effect and Compton scatter when x-ray photon energy range is below 200 keV. In DECT reconstructions, generally two basis materials can be selected for decomposition, for example, water and iodine, bone and soft tissue. There are two kinds of dual energy material decomposition methods in the literature: pre-decomposition and post-decomposition. Pre-decomposition utilizes the spectral information of low and high energy thus is considered as accurate and exact while post-decomposition ignores the nonlinear property of polychromatic spectrums thus is approximate. If a constraint of volumetric fraction is applied, three-material decomposition with dual-energy or two-material decomposition with single-energy becomes possible.

There are three most common configurations for the dual-energy data acquisition: dual-source, sandwiched dual-layer/photon-counting detectors and fast kVp-switching. A dual-source DECT has two sets of X-ray tube and detector. Each of them with individual filtering operates on kVp energy. The drawback of this setup is scatter interference and high system cost. A dual-layer DECT has one X-ray tube and two overlapping detector layers. Its front layer primarily absorbs the lower energy photons while the under layer detects the rest higher energy photons behind a metal filter. The cost of this setup is still higher than a conventional single energy CT. For a fast kVp-switching DECT, its X-ray tube voltage alternates rapidly between low and high kVp and the detector acquires two projections at each view. In such a system, the X-ray tube voltage must be switched at least about 2 kHz to meet modern CT imaging requirement. Very recently, a dual-energy CT using slow kVp switching acquisition and a prior image constrained compressed sensing is proposed which reduces the voltage switch frequency to one eighth of the original rate and produces comparable image quality[1].

In this work, we present a new dual-energy scheme suitable for conventional single energy CT systems with few hardware modifications. In the proposed scheme, a circular trajectory is separated into two semi-circles. For each semi-circle, the X-ray tube operates at low or high voltage respectively, i.e., $[0, \pi]$ for low energy and $[\pi, 2\pi]$ for high energy. Therefore, both the low and high energy projection is of limited angular coverage. In general, this approach can be naturally extended to tri-energy or even more multi-energy. We refer this scheme as segmented trajectory multi-energy CT (SegMECT). Compared with a fast kVp-switching DECT, the voltage change frequency of SegMECT is reduced to Hz magnitude. A prior image based iterative reconstruction algorithm is proposed to solve this limited-angle MECT problem, which is described in detail in the following sections.

II. Method

A. Multi-Energy Material Decomposition

The physical model of dual-energy CT is based on polychromatic X-ray imaging. The detector output signal I can be expressed as:

$$I_{low} = \int_E S_{low}(E) D(E) W(E) e^{-\int_{\mathbf{r}} \mu(E,\mathbf{r}) d\mathbf{r}} dE$$
$$I_{high} = \int_E S_{high}(E) D(E) W(E) e^{-\int_{\mathbf{r}} \mu(E,\mathbf{r}) d\mathbf{r}} dE \quad (1)$$

This word is supported by the National Natural Science Foundation of China on project NO. 11275104.

The authors are with Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education and the Department of Engineering Physics, Tsinghua University, Beijing 100084 China.

Yuxiang Xing is the corresponding author: phone: 8610-62782510, e-mail:xingyx@mail.tsinghua.edu.cn.

where $S(E)$ denotes the low and high energy spectrum of the X-ray tube, $D(E)$ the detector response, $W(E)$ the energy weighting (always proportional to energy), and $\mu(E,\mathbf{r})$ the spatial distribution of linear attenuation coefficient. According to the material decomposition method, the linear attenuation coefficient can be expressed as linear combination of two basis material:

$$\mu(E,\mathbf{r}) = b_1(\mathbf{r})\mu_1(E) + b_2(\mathbf{r})\mu_2(E) \tag{2}$$

If the attenuation images of dual-energy $\mu_{low}(\mathbf{r})$ and $\mu_{high}(\mathbf{r})$ are reconstructed, the basis coefficients distribution $b_1(\mathbf{r})$ and $b_2(\mathbf{r})$ can be calculated by solving the following linear equations:

$$\begin{pmatrix} \mu_{1,low} & \mu_{2,low} \\ \mu_{1,high} & \mu_{2,high} \end{pmatrix} \begin{pmatrix} b_1(\mathbf{r}) \\ b_2(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} \mu_{low}(\mathbf{r}) \\ \mu_{high}(\mathbf{r}) \end{pmatrix} \tag{3}$$

The coefficient matrix is usually generated via a calibration scan of the two basis materials under the same dual-energy spectra. Compared with (2), the energy variable is neglected in (3), thus, the finally decomposition results may suffer from beam-hardening or metal artifacts. When it comes to multi-energy, suppose there are M different spectra and N basis materials (with the constraint $M \geq N$), the equations become:

$$\begin{pmatrix} \mu_{1,E_1} & \cdots & \mu_{N,E_1} \\ \vdots & \ddots & \vdots \\ \mu_{1,E_M} & \cdots & \mu_{N,E_M} \end{pmatrix} \begin{pmatrix} b_1(\mathbf{r}) \\ \vdots \\ b_N(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} \mu_{E_1}(\mathbf{r}) \\ \vdots \\ \mu_{E_M}(\mathbf{r}) \end{pmatrix} \tag{4}$$

B. Multi-Energy Iterative Reconstruction Algorithm

We show in Fig.1 the multi-energy CT scan scheme with three energy spectrums with angular segments of $2\pi/3$. In general, suppose there are M spectrums, and the scan range of jth spectrum would be $[2\pi(j-1)/N,\ 2\pi/N\ ]$ (commonly N<6). Denote the collected limited-angle projection data (with logarithmic transform) as $\mathbf{p}_{E_j}$, the $2\pi/N$ limited-angle projection matrix as $\mathbf{H}_{E_j}$ and the corresponding attenuation image as $\mathbf{f}_{E_j}$. The multi-energy reconstruction problem can be formulated as:

$$\mathbf{H}_{E_j}\mathbf{f}_{E_j} = \mathbf{p}_{E_j} \quad j = 1,\cdots,M \tag{5}$$

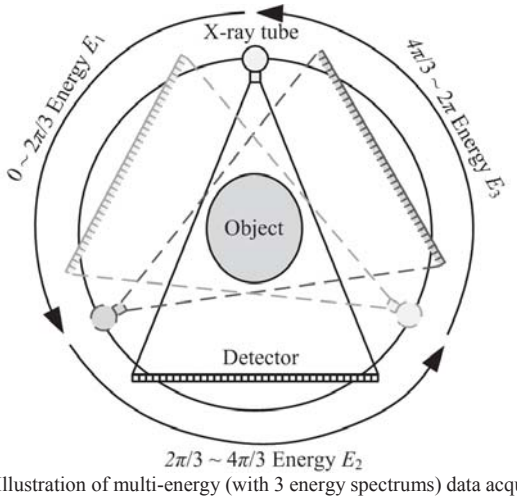


Fig. 1. Illustration of multi-energy (with 3 energy spectrums) data acquisition scheme.

The above problem is likely to be rather ill-posed. In order to get better image quality, additional constraint is involved. Total variation minimization is a kind of widely used constraints. Among them, the anisotropic total variation (ATV) is a recent proposed constrained minimization reconstruction framework for a limited-angle CT problem[2]. Instead of 2D isotropic TV minimization, the ATV method solves a multi-1D TV minimization along a series of specific directions:

$$\begin{aligned} &\min \sum_t k_t \left\| \nabla_{\alpha_t} \mathbf{f}_{E_j} \right\|_1, \quad \text{s.t.} \ \left\| \mathbf{H}_{E_j}\mathbf{f}_{E_j} - \mathbf{p}_{E_j} \right\|_2 \leq \varepsilon_{E_j} \\ &\sum_t k_t^2 = 1 \end{aligned} \tag{6}$$

where k_t is the normalization weighting factor and $\nabla_{\alpha_t} = (\cos\alpha_t, \sin\alpha_t)\bullet\nabla$ is the directional gradient operator. A practical discrete implementation can be expressed as:

$$\begin{aligned} \left\| \nabla_{\alpha_t} \mathbf{f}_{E_j} \right\|_1 &= \left\| (\cos\alpha_t, \sin\alpha_t)\bullet\nabla \mathbf{f}_{E_j} \right\|_1 \\ &= \left\| \cos\alpha_t \bullet \dot{\mathbf{f}}_{E_j,x} + \sin\alpha_t \bullet \dot{\mathbf{f}}_{E_j,y} \right\|_1 \\ &\cong \sum \left| \cos\alpha_t \left(f_{x+1,y} - f_{x,y} \right) + \sin\alpha_t \left(f_{x,y+1} - f_{x,y} \right) \right| \end{aligned} \tag{7}$$

Although the intensity of attenuation images at different energy differs from each other, there is strong structural similarity between them. Based on this similarity, a prior image can be generated to help to constrain the iteration progress. For the energy E_j, first the complete projection data is generated with intensity equalization:

$$\begin{aligned} \tilde{\mathbf{p}}_{E_j} &= \left(w_1 \mathbf{p}_{E_1}^T, \cdots, w_M \mathbf{p}_{E_M}^T \right)^T \\ w_i &= \frac{\left\| \mathbf{p}_{E_j} \right\|_1}{\left\| \mathbf{p}_{E_i} \right\|_1} \end{aligned} \tag{8}$$

Then a standard FBP reconstruction is carried out and the prior image $\tilde{\mathbf{f}}_{E_j}$ is obtained. In the PICCS algorithm [1] the target image is reconstructed by minimize the following objective function:

$$\begin{aligned} &\min_{\mathbf{f}} \left[(1-\alpha) \left\| \Psi_1(\mathbf{f}) \right\|_1 + \alpha \left\| \Psi_2\left(\mathbf{f} - \mathbf{f}_p \right) \right\|_1 \right] \\ &s.t. \quad \mathbf{Hf} = \mathbf{p} \end{aligned} \tag{9}$$

The functions Ψ_1 and Ψ_2 are some sparse transforms, typically the gradient norm. This constrained minimization problem is usually solved by alternative minimizing between data fidelity and TV norm. In this work, we use ATV for Ψ_1. Moreover, rather than using the TV norm of the difference image $\mathbf{f} - \mathbf{f}_p$, we apply TV norm of quotient image to be the second objective term[3]:

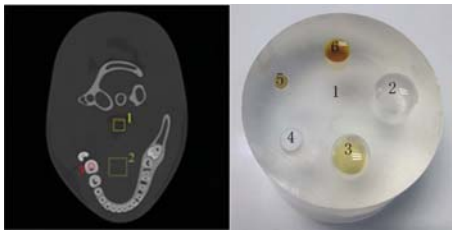


Fig. 2. Phantoms and their ROI used in this work. Left: Digital dental phantom for numerical simulation. Right: Physical cylindrical phantom for the practical experiment.

TABLE I
MATERIAL DEFINITION OF PHYSICAL PHANTOM

Index	Material
1	PMMA
2	absolute alcohol
3	20% iodine solution[a], 80% alcohol
4	Teflon
5	polyimide
6	100% iodine solution

[a]Anerdian, a commonly used medical disinfectant. Ingredients: 0.2% iodine, 70% alcohol.

$$TV\left(\mathrm{Diag}^{-1}\left(\mathbf{f}_p\right)\bullet\mathbf{f}\right)=\left\|\nabla\left(\frac{\mathbf{f}}{\mathbf{f}_p}\right)\right\|_1 \tag{10}$$

The overall objective function is:

$$\min_{\mathbf{f}_{E_j}}\left[(1-\alpha)\sum_t k_t\left\|\nabla_{\alpha_t}\mathbf{f}_{E_j}\right\|_1+\alpha\left\|\nabla\frac{\mathbf{f}_{E_j}}{\tilde{\mathbf{f}}_{Ej}}\right\|_1\right] \tag{11}$$
$$s.t.\quad \left\|\mathbf{H}_{E_j}-\mathbf{f}_{E_j}\right\|_2\le\varepsilon_j$$

After all the attenuation images are reconstructed, the post-process material decomposition is carried out by a simple matrix inversion on (4) to obtain the basis coefficients.

III. Simulation and Experiments

A. Polychromatic Data Simulation

A dental phantom (in Fig.2 left) based on B-spline curve was constructed with about 10 materials, such as bone, muscle, adipose, dentine, enamel, etc. The elements composition and concentration were collected in [4]. Linear attenuation coefficients were generated by the XCOM program[5].

X-ray spectrums were calculated with the SpekCalc software[6]. The spectrums used for simulation are shown in Fig.3. The 90 and 120 kVp energies with 2.0 mm of Cu filtration were configured for dual-energy simulation. The energy-integrated spectral projection were calculated by (1). For dual-energy CT, the dental phantom was sized to 38 cm. A detector with 960 channels was placed 75 cm always from the X-ray source. A total of 360 view angles were acquired per rotation. The initial photon numbers for each detector bin are 4.76e5 of 120 kVp and 7.27e5 of 90 kVp, which yield about equivalent noise level of the dual-energy projection data.

B. Experimental Measurements with a Physical Phantom

As shown in Fig. 2 right, we manufactured a PMMA cylinder of 5 cm diameter with five cylindrical inclusions of descending diameter filled with Teflon rod, polyimide rod, iodine solution and alcohol listed in Table. I. The projection data was acquired on a circular cone-beam CT with a flat panel detector. Three complete measurements were taken for 60, 90, 120 kVp energies respectively. The number of views per rotation is 360. One third of each data made up the limited-angle data for SegMECT of 3 segments.

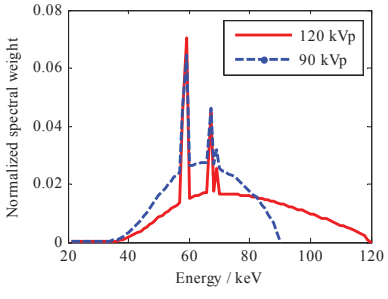


Fig. 3. Synthetic spectra (120 and 90 kVp) for dual-energy CT simulation.

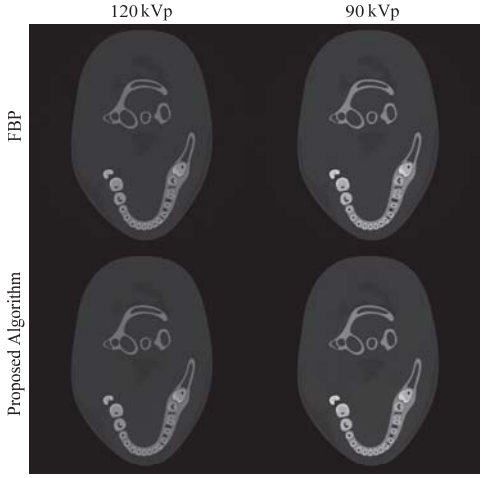


Fig. 4. Reconstruction results of dual-energy CTs. First row: results from standard FBP algorithm with complete data. Second row: results from our proposed algorithm with the limited-angle data from a SegMECT of 2 segments.

IV. Results and Discussion

A. Numerical study

The minimization problem in (11) was numerically solved by the ASD-POCS algorithm[7]. Since the total variation of the quotient image previously defined in (10) may be impacted by the noise in the prior image, noise suppression was performed to get a clear prior image. When calculating the steepest descent direction of the total variation, pixels representing air in the prior image were ignored to avoid the division by zero. The threshold of air and object was set to be 0.035. We show final results in Fig. 4. Compared with the FBP results, the proposed algorithm gets rid of noise and streaking artifacts. Besides, the edges and structures are well preserved because of the prior information. The noise standard derivation of the proposed

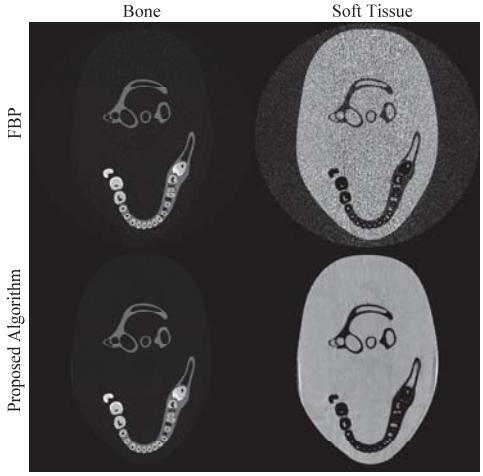


Fig. 5. Dual-energy decomposition coefficients of basis materials resulted from the attenuation images in Fig. 4.

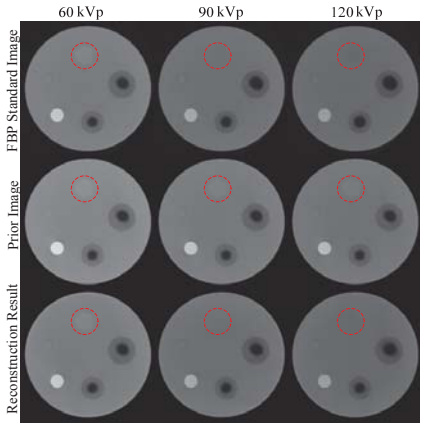


Fig. 6. Experimental results of attenuation of 60, 90, 120 kVp. All images are displayed at [-0.05, 0.5] cm^{-1}. First row: standard images of FBP for reference. Second row: prior images of three energies. Third row: results of the proposed algorithm. Bottom row: error images between the results and reference images.

algorithm is an order of magnitude less than that of FBP results. The decomposition images of bone and soft tissue are presented in Fig. 5. In the FBP results, the noise is amplified by the inversion coefficient matrix in (3) since its entries' absolute values are much larger than 1. While in the proposed results, the absence of noise in dual-energy attenuation images maintains a lower noise level in the decomposed images. However, slight limited-angle artifacts appear in the bilateral boundaries of the phantom, which are more than likely to be suppressed by carefully tuned parameters.

B. Experimental Results

In order to make a fair comparison, we denoised the acquired projection data in the projection domain. Meanwhile, only air threshold of 0.03 was applied in the iteration. The attenuation images of three energies are shown in Fig. 6. Images in the first row are the standard reference images of FBP with complete data. Images in the second and third rows are the prior images and final reconstruction results of the proposed method. The results of our proposed method are free of limited-angle artifacts. The SNR and CNR of the proposed method are close to those of FBP algorithm. The region of interest (ROI) circled by dashed red line refer to 100% iodine solution. The expectation values of it are greater than, equal to and less than the background of PMMA for 60, 90, 120 kVp in the true images. However, the ROI areas in the prior images are always greater than the background for all three energies. Even so, this improper prior of density can be corrected during the iteration by the data fidelity term. The error images in the last row demonstrate the effectiveness of our method. The decomposition images of PMMA, absolute alcohol and iodine solution are presented in Fig. 7. The proposed results also have a better SNR than the FBP results. The emphasizing of iodine component in the decomposed images implies the ability of material separation in multi-energy CT when contrast agent gets involved in.

C. Discussion

According to the numerical and experimental studies we carried out, our proposed method proves to be capable of dealing with the limited-angle reconstruction problem derived by the slow-switch multi-energy CT scheme SegMECT. Because of the X-ray polychrome, the line-integral model (5) no longer establishes. Thus, the inconsistent error would spread along the direction of missing data. The prior image generated by the combination of multi-energy data can provide most of the structure and detail information. Noise suppression and other correction such as metal artifact reduction are essential to be carried out so that better image quality can be obtained.

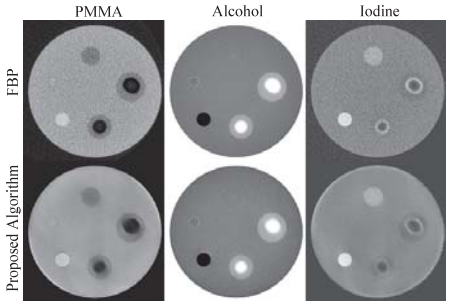


Fig. 7. Coefficients of three-energy basis material decomposition.

V. Conclusion

In this work, we proposed a new SegMECT to enable multi-energy scan on a conventional CT system with little additional cost. A constrained objective function based on the TV norm of a quotient image and ATV was established. The post-decomposition of sbasis material was utilized to obtain the material specific images. Both numerical simulation and practical experiments demonstrated the effectiveness of the proposed algorithm.

Although the angular coverage of each spectrum is equal here, there might be alternations, e.g., a short scan for low energy and the rest angular range for high energy. Thus, the prior image may be accurately reconstructed without any artifact, which could be more helpful in solving the limited-angle problem. This work can be easily extended to volumetric + spectral imaging.

References

[1] T. P. Szczykutowicz and G.-H. Chen, "Dual energy CT using slow kVp switching acquisition and prior image constrained compressed sensing," *Physics in Medicine and Biology,* vol. 55, p. 6411, 2010.

[2] X. Jin*, et al.*, "Anisotropic total variation minimization method for limited-angle CT reconstruction," in *Developments in X-Ray Tomography VIII*, San Diego, California, USA 2012, pp. 85061C-85061C.

[3] Y. Xing and P. Zheng, "A restoration method for incomplete data in DECT," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, 2011, pp. 4306-4310.

[4] Y. H. Na, "Deformable Adult Human Phantoms for Radiation Protection Dosimetry: Methods for Adjusting Body and Organ Sizes to Match Population-Based Percentile Data," PhD, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, 2009.

[5] M. J. Berger*, et al.* (1998). *XCOM: Photon Cross Sections Database*. Available: http://www.nist.gov/pml/data/xcom/index.cfm

[6] G. Poludniowski*, et al.*, "SpekCalc : a program to calculate photon spectra from tungsten anode x-ray tubes," *Physics in Medicine and Biology,* vol. 54, p. N433, 2009.

[7] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology,* vol. 53, p. 4777, 2008.

Combined Respiratory and Rigid-Body Motion Compensation in Gated Cardiac Perfusion SPECT using a Visual Tracking System

P. Hendrik Pretorius, Michael A. King, Karen L. Johnson, and Michael O'Connor.

***Abstract*-- We report on early results of our combined respiratory and rigid-body motion compensation strategy in 17 gated cardiac perfusion single photon emission computed tomography (SPECT) patient studies, employing a visual tracking system (VTS). Patient volunteers with written consent were imaged employing list-mode acquisitions during gated stress Tc-99m sestamibi perfusion SPECT on a BrightView SPECT/CT (Philips, Cleveland, OH). Motion tracking was performed using 5 near infrared Vicon cameras in combination with 6 or 7 retro-reflective markers. Processing steps included, down sampling VTS positional data from 30 Hz to 10 Hz (100 ms), synchronizing down sampled VTS data with 100 ms SPECT frames, separating rigid-body and respiratory motion, estimating 6-DOF rigid-body motion using singular value decomposition (SVD), amplitude binning 100 ms non-gated SPECT frames into respiratory projection sets, amplitude binning individual 8 ECG gates with frame lengths varying between ~70 ms and 112 ms into 8 respiratory projection sets, reconstructing the non-gated SPECT respiratory projection sets with rigid-body motion compensation, estimating cardiac respiratory motion employing intensity based registration, combining rigid-body and respiratory motion estimates, and reconstructing non-gated and gated amplitude binned SPECT projections with combined compensation. We determined in the first 17 patient acquisitions that combined respiratory and rigid-body motion compensation is possible for gated perfusion SPECT without noise degradation compared to standard gated perfusion SPECT. Although respiratory estimates in the 16 patients were of small to moderate in extent (1-15.75 mm in the axial direction), improved cardiac motion clarity was observed due to the reduction in respiratory blurring. It is envisaged that respiratory compensation will be most beneficial for patients with small to moderate sized hearts with respiratory amplitudes larger than 10 mm.**

I. Introduction

Since the early days of cardiac perfusion single photon emission computed tomography (SPECT) it has been known that respiratory motion reduced image quality [1]. More than a decade ago we demonstrated this degradation in a digital cardiac perfusion phantom study [2] and others have confirmed our findings [3-5]. Since that time we have implemented various rigid-body and respiratory motion compensation strategies [6-13], either separately [8-10, 13] or in some combination [6, 7, 11]. We are now able to correct for both these corrections with minimal initial user interaction employing a visual tracking system (VTS) from Vicon Motion Systems, Inc. (Lake Forest, CA) for motion tracking.

In this study we report on early results of our combined respiratory and rigid-body motion compensation strategy in 17 gated cardiac perfusion single photon emission computed tomography (SPECT) patient studies, employing the visual tracking system (VTS).

Manuscript received February 14, 2013. This work was supported by the National Institute of Biomedical Imaging and Bioengineering grant R01-EB001457 and a research grant from Philips Healthcare. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or Philips Healthcare. P. Hendrik Pretorius (email: Hendrik.Pretorius@umassmed.edu), Michael A. King, Karen L. Johnson, and Michael O'Connor are from the University of Massachusetts Medical School Worcester, MA 01655 USA.

II. Methods and Materials

A. Patient Acquisitions

Patient volunteers with written consent were asked to breathe normally during a list-mode gated stress Tc-99m sestamibi SPECT perfusion acquisition as part of a rest-stress perfusion exam. The acquisition was performed using the standard clinical protocol with 64 projections acquired through 180 degrees employing a two-headed BrightView SPECT/CT scintillation camera (Philips, Cleveland, OH), which rotated with 2.825-degree steps during emission acquisition. The pixel size was 0.4662 cm in a 128x128 acquisition matrix. Cone-beam CT acquisitions were acquired through 360 degrees in 0.83-degree steps. The patients were monitored by the VTS throughout the whole study with markers positioned as detailed below.

B. Visual Tracking System Acquisitions

Motion tracking was accomplished by employing 5 near-infrared Vicon cameras in combination with the retro-reflective markers positioned on the patient volunteers. The placement of the retro-reflective markers is shown in fig 1, with 2 on the right side of the chest and 3 on the left. In the current study one retro-reflective marker was placed on the abdomen to track abdominal respiration and an additional retro-reflective sphere was placed on the lower right edge of the ribcage in a subset of patients to add more stability to the singular value decomposition (SVD) estimation of motion (fig 1).

Two different ways were used to accurately synchronize the VTS and SPECT system. In the first method a repeating digital pulse, with a period of 1s, generated by the VTS system was inserted into the list during list mode acquisition starting 240 sec after commencement of visual tracking. At the same time ECG gate signals were also inserted into the list by the standard Philips hardware. To accomplish this without disrupting the normal acquisition flow, the two signals are inserted via two different digital ports. Activation of the second port to acquire synchronization signals required intervention by Philips personnel and was disabled whenever the acquisition computer was rebooted. Second, we developed a method to insert a unique sequence of short pulses 25 sec after visual tracking starts, but before the start of the emission acquisition (during the CT acquisition). This sequence of pulses was multiplexed with the ECG signal. This enabled us to synchronize the VTS with the

SPECT system acquisition without the need to enable the second port.

Fig 1: The placement of the retro-reflective markers on patient volunteers.

C. Rigid-Body and Respiratory Motion Estimation

Processing was fully automatic except for the placement of a predefined ellipsoidal region of interest (ROI) around the heart using an initial reconstruction. Processing steps were 1) reformatting list-mode data into 64 projections as well as 100 ms frames, 2) down sampling the 30 Hz VTS positional measurements to 10 Hz (100 ms) and temporal synchronization with 100 ms SPECT frames, 3) separating rigid-body and respiratory motion components of the motion tracking signals from the VTS and estimating 6 degrees-of freedom (DOF) rigid-body motion as described in [11] using SVD, 4) amplitude binning the 100 ms frames into an odd number (N) of respiratory projection data sets for both non-gated and gated SPECT reconstruction, 5) scaling and reconstruction with rigid-body motion compensation N-1 projection sets, with the center or reference bin excluded [7], 6) scaling and reconstructing with rigid-body motion compensation N-1 unique reference projection sets such that each reference projection set has the same number of non-zero projections as the individual projection sets in previous step [7], 7) estimating respiratory motion employing the ellipsoidal ROI to isolate the heart on the reconstructed reference bin in combination with an intensity based estimation method, and 8) combing rigid body and respiratory motion estimates in a final single non-gated reconstruction as well as 8 gated reconstructions employing all the acquired data.

D. Reconstructions

An ordered subsets expectation maximization (OSEM) algorithm incorporating a 3-dimensional Gaussian rotator [6] was used for all reconstructions. During respiratory estimation reconstructions, 16 projections per subset were employed. The reason for the large number of projections per subset during respiratory motion estimation reconstruction was to minimize reconstruction artifacts when incomplete data were present (amplitude bins with little or no counts at some projection angles were discarded), but still obtain some measure of speed up compared to our first efforts in [9] where maximum likelihood expectation maximization (MLEM) were used. Only 3 iterations of OSEM were employed (~12 MLEM equivalent iterations as in ref. 9). The final combined compensation for rigid-body and respiratory motion were implemented by reconstructing each projection as N sub-projections (equal to the number of respiratory amplitude bins) sequentially before stepping to the next projection in the subset. Each of the sub-projections was weighted with its fractional contribution to the full projection. Rigid-body motion compensation reconstructions (without respiratory motion compensation) were also performed for comparison. The gated reconstructions with and without respiratory motion compensation were reconstructed with 8 projections per subset All reconstructions included attenuation, non-circular resolution, and triple-energy window scatter compensation.

E. Image Analysis and Evaluation

These final reconstructions were Gaussian post-reconstruction filtered, and reoriented to short axis, horizontal long axis, and vertical long axis slices. Visual inspection of the gated data and polar map quantification of the non-gated data were used for evaluation. It is planned to further analyze the gated data by determining the change in ejection fraction between standard and respiratory motion corrected data sets.

III. Results and Discussion

We determined from the first 17 patient acquisitions that combined respiratory and rigid body motion compensation is possible for gated perfusion SPECT with little noise degradation compared to standard gated perfusion SPECT. As expected, little or no body motion was detected as the patients were asked not to move. Although respiratory estimates in the 17 patients were of moderate extent (<1-15.75 mm in the axial direction), improved cardiac motion clarity was observed. For larger respiratory motion, compensated gated SPECT shows an increase in inferior and anterior uptake as well as improved base to apex motion for the respiratory compensated data. Such a patient example is given in fig's 2-5. In our patient set, the patient with the smallest heart (visually determined using size and number of slices) had an estimated 8.9 mm respiratory motion. Non-gated polar maps of this patient show a slight improvement in inferior and anterior uptake after respiratory compensation (fig. 6, left) compared to the standard reconstruction (fig. 6, center), while the difference image clearly demonstrate the improvement. A larger dynamic range in counts for respiratory motion compensated gated polar maps (bottom row, fig. 7) are notable compared to polar maps using the standard reconstruction (fig. 7, top row). A larger dynamic range in counts is indicative of improved function.

Based on our studies thus far we hypothesize that respiratory compensation will be most beneficial for patients with small to moderate sized hearts with respiratory amplitudes larger than 10 mm. In this study 3 of the patients have respiratory estimates in excess of 10 mm. In a previous study of 32 patients, 11 of the patients had respiratory estimates larger than 10 mm [11].

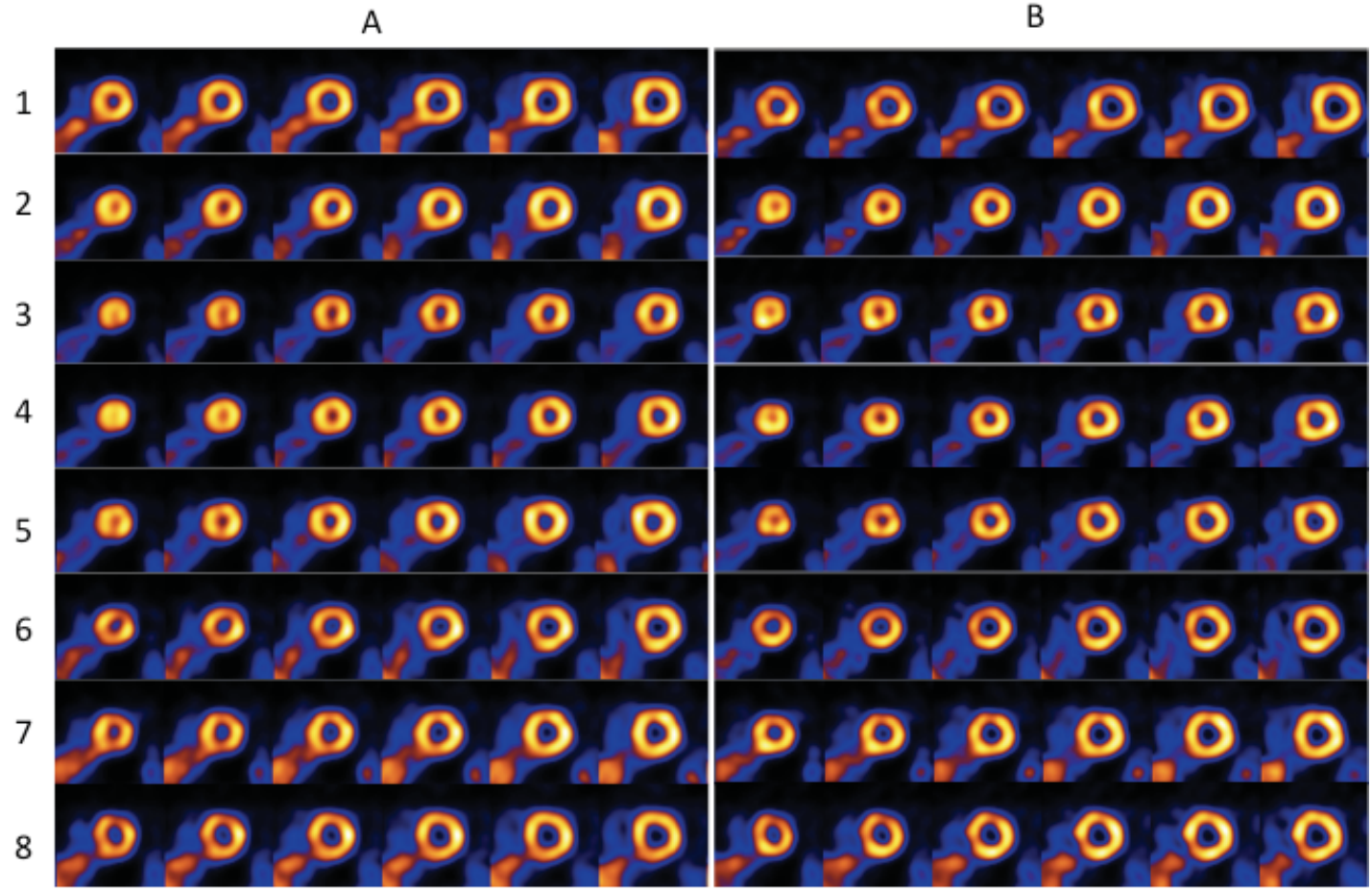


Figure 2: Short axis slices of a male volunteer with approximately 15.75 mm respiratory motion. Respiratory motion corrected data are given in column B while the standard gated reconstructions are displayed in column A. Eight gates (1-8) are given, each displayed to its own maximum. The effect of respiratory motion is clearly seen in both the slightly elongated shape of the cavity and the blurring of the heart wall. Note the inferior cooling that persist throughout in column A compare to column B. The liver counts in the background are helpful to judge the change in myocardial counts between gates. Note that the background counts appear to be less in column B, indicating higher count levels in the heart wall.

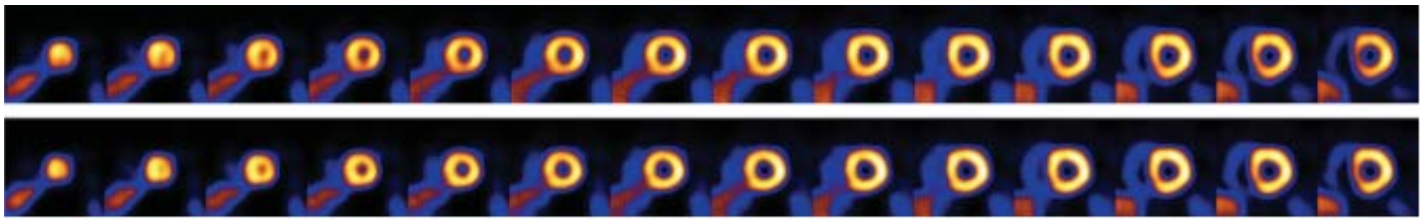


Fig 3. Short-axis slices in absence of ECG GATING of the same male patient volunteer as in fig. 2. Notice the improved blood pool and wall visualization in bottom row with respiratory motion correction compared to the top row. The cooling in the inferior wall when respiratory compensation is absent is also visible (top row).

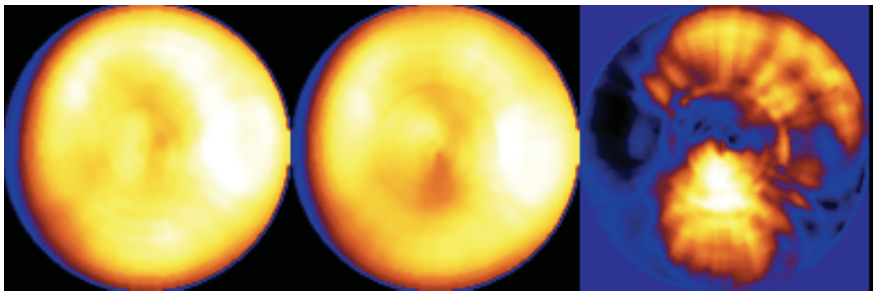


Fig 4: Polar maps of the short axis slices displayed in fig. 3. On the left the polar map generated from the data with respiratory compensation, in the center data without respiratory motion compensation, and a difference polar map on the right. Note the differences in the inferior and anterior portions of the polar maps indicating the reduction in the artifactual cooling in these walls due to the superior / inferior motion of the heart with respiration.

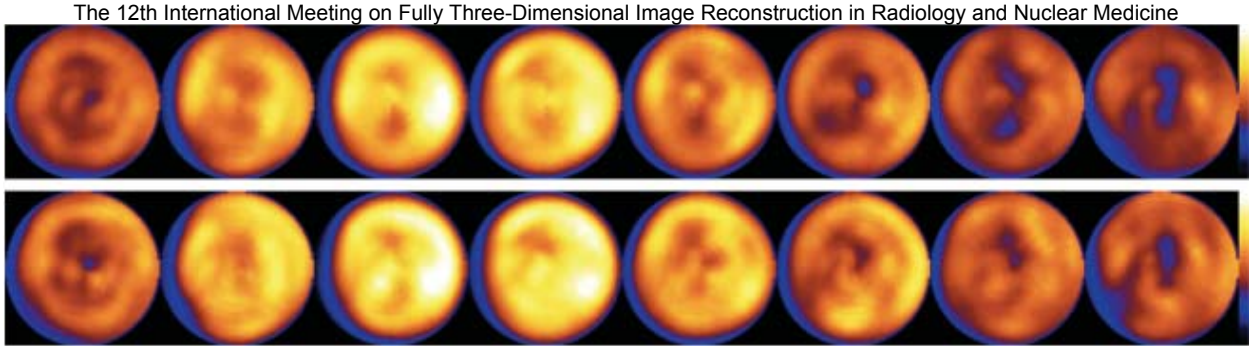


Fig 5: Polar maps of the short axis slices displayed in fig. 2. The top row depicts the polar maps of the standard gated reconstructions while the bottom row are the polar maps generated using the respiratory compensated data. Both rows of data were normalized to the maximum count in the row (polar map with highest counts in the row). The variation in thickness during contraction and relaxation are therefore shown as an increase and decrease in counts (see color bar). The top row again shows the artifactual cooling in the inferior and anterior portions of the polar map due to respiratory motion.

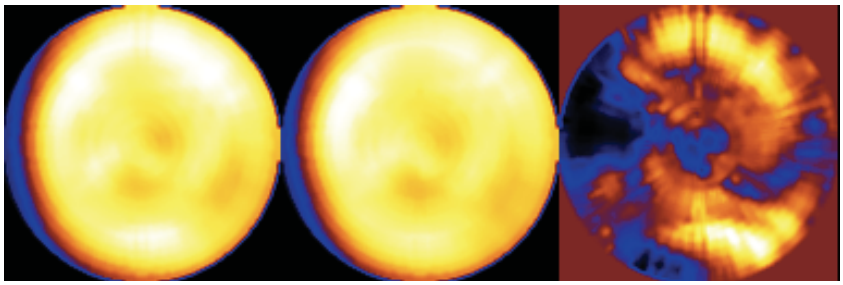


Fig 6: Non-gated polar maps for the patient with the smallest heart. On the left the polar map with respiratory motion compensation, in the center the polar map without respiratory motion compensation, and a difference image on the right. Even with only 8.9 mm respiratory motion, the differences in uptake in the inferior and anterior portions of the polar maps are visible.

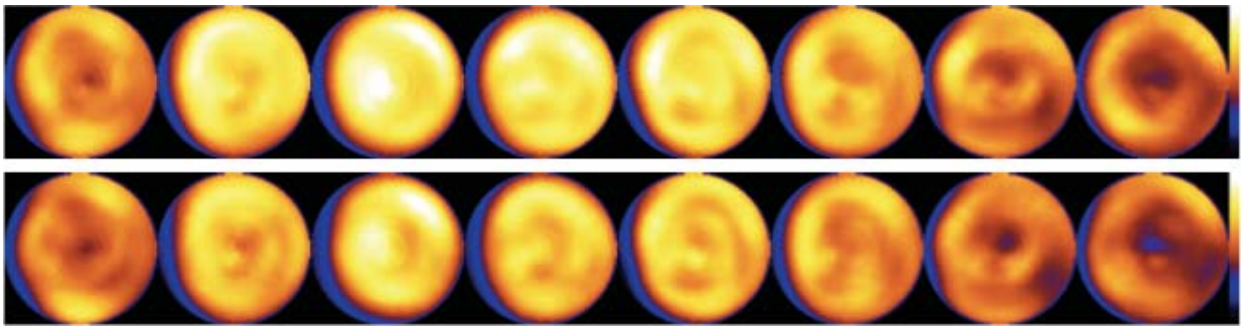


Fig 7: Gated polar maps for the patient with the smallest heart. The top row depicts the polar maps of the standard gated reconstructions while the bottom row are the polar maps generated using the respiratory compensated data. The dynamic range in counts, an indication of improved function is visible in the bottom row.

REFERENCES

[1] M.M. Ter-Pogossian, S.R. Bergman, and B.E. Sobel, "Influence of cardiac and respiratory motion on tomographic reconstructions of the heart: Implications for quantitative nuclear cardiology," *J. Comp. Assist. Tomog.*, vol 6, pp. 1148-1155, 1982.

[2] P.H. Pretorius and M.A. King, "A study of possible causes of artifactual decreases in the left ventricular apex with SPECT cardiac perfusion imaging," *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 1016-1023, 1999.

[3] B.M.W. Tsui, W.P. Segars, D.S. Lalush, "Effects of upward creep and respiratory motion in myocardial SPECT," *IEEE Trans. Nucl. Sci.*, vol. 47, pp. 1192–1195, 2000.

[4] K. Cho, S. Kumiata, S. Okada, T. Kumazaki, "Development of respiratory gated myocardial SPECT system," *J Nucl Cardiol.*, vol. 6, pp. 20-28, 1999.

[5] G. Kovalski, O. Israel, Z. Keidar, A. Frenkel, J. Sachs and H. Azhari, "Correction of heart motion due to respiration in clinical myocardial perfusion SPECT scans using respiratory gating". *J. Nuc Med*, vol. 48, pp. 630-636, 2007.

[6] B. Feng, J. Dey, P.H. Pretorius, R.D. Beach, J.E. McNamara, M.S. Smyczynski K. Johnson, and M.A. King, "Compensation for rigid-body patient motion during reconstruction and respiratory motion post-reconstruction in phase-binned slices," Proceedings of 2006 IEEE Medical Imaging Conference, M06-290, 2103-2106, 2006.

[7] R. Beach, H. Depold, G. Boening, P.P. Bruyant, B. Feng, H. Gifford, M. Gennert, N. Suman, and M.A. King, "An adaptive neural network approach to decomposition of patient stereo-infrared tracking motion data during cardiac SPECT imaging using asymmetric median filters," *IEEE Trans. Nucl. Sci.*, vol. 54, pp. 130-139, 2007.

[8] J.E. McNamara, P.H. Pretorius, K. Johnson, J. Mitra, J. Dey, M.A. Gennert, and M.A. King, "A flexible multi-camera visual-tracking system for detecting and correcting motion-induced artifacts in cardiac SPECT slices," *Med. Phys.*, vol. 36, pp. 1913-1923, 2009.

[9] J. Dey, W.P. Segars, P.H. Pretorius, R.P. Walvick, P.P. Bruyant, S Dahlberg, and M.A. King, "Estimation and correction of cardiac respiratory motion in SPECT in the presence of limited angle effects due to irregular respiration," *Med. Phys.*, vol. 37, pp. 6453-6465, 2010.

[10] P.H. Pretorius, K.L. Johnson, J.M. Mukherjee, and M.A. King, "Accuracy of motion estimation and correction in cardiac SPECT with a clinical visual tracking system (VTS): a phantom study," *J. Nucl. Med.*, 52, 47P, 2011

[11] P.H. Pretorius, M.A. King, K.L. Johnson, J.M. Mukherjee, J. Dey, A. Konik, "Combined Respiratory and Rigid Body Motion Compensation in Cardiac Perfusion SPECT Using a Visual Tracking System," Proceedings of 2011 IEEE Medical Imaging Conference, 2768-2773, 2011.

[12] PH Pretorius, KL Johnson, MA King, "Alignment of attenuation and emission data using a visual tracking system," J Nucl Med, 53, P146, 2012.

[13] J. M. Mukherjee, J. E. Mcnamara, K. L. Johnson, J. Dey, M. A. King, "Estimation of Rigid-body and Respiratory Motion of the Heart for SPECT Motion Correction," *IEEE Trans. Nucl. Sci*., vol. 56, pp. 147-155, 2009.

Super-Resolution PET Using Stepping of a Deliberately Misaligned Bed

Scott D. Metzler, *Senior Member, IEEE,* Yusheng Li, *Member, IEEE,*
Joel S. Karp, *Fellow, IEEE,* and Samuel Matej, *Senior Member, IEEE*

***Abstract*—A typical PET scanner acquires data from a patient for several bed positions, with some overlap between bed positions. When the bed is aligned with the scanners axial direction, the sampling for portions of the body acquired over more than one bed position does not change. When a small rotation is introduced, combined with vertical stepping and near-continuous axial translation, multiple data sets, each with a unique position relative to the scanner's lines of response, are acquired. These super-resolution data sets can be combined in reconstruction to potentially yield a higher resolution reconstruction with reduced artifacts.**

***Index Terms*—PET, sampling, super-sampling, wobbling.**

I. Introduction

IT has been shown in clinical and preclinical systems that super-sampling (e.g., controlled wobbling of the bed) leads to improved reconstruction uniformity with a reduction in aliasing artifacts [1]–[9]. In fact, early PET scanners used wobbling techniques to improve reconstruction quality that was degraded from undersampling. However, the introduction of block detectors enabled the evolution to smaller crystal sizes and improved angular and radial sampling [5], [7], [9]. Their use made whole-body scanners commercially viable [5], [7], [9] and reduced the impact of crystal size on overall resolution compared to other effects, such as acolinearity and the block-detector effect. This had the effect of improving sampling compared to resolution, reducing the impact of wobbling on reconstruction quality. Thus, when producing the first whole-body PET scanners, manufacturers dropped wobbling as mechanically cumbersome, unneeded for reconstruction, and an extra expense.

Two important characteristics of scanners have changed, making it important to reconsider super-sampling: (i) scanners are now 3D, providing more counts, allowing accurate reconstruction of finer structures [10]; and (ii) iterative algorithms with resolution modeling have changed again the relationship between reconstruction resolution and sampling, making it important to boost the number of samples. Consequently, it is possible that super-sampling in modern PET scanners will allow resolution improvement with iterative algorithms using resolution-recovery, beyond what those algorithms can achieve without super-sampling.

Clinical acquisitions often use multiple bed positions, where the patient is scanned for a few minutes and then shifted by a fraction of the axial field of view (FOV) along the scanner's axial direction. After the axial shift, the voxels remaining in the FOV have the same position relative to the lines of response (LORs), at least in the transverse directions (i.e., the horizontal and vertical shifts are zero) and usually in the axial direction; this wastes an opportunity to improve the sampling at essentially no cost. We exploit this opportunity by introducing a small angle between the bed motion and the scanner's axial direction to super-sample in the horizontal direction. In conjunction, we use vertical bed movement to super-sample in the vertical direction. Axial super-sampling comes from axial bed shifts by a non-integral number of crystal pitches. An additional benefit of using nearly continuous axial motion is axial uniformity of sensitivity (noise).

S.D. Metzler, Y. Li, J.S. Karp, and S. Matej are with the Department of Radiology, The University of Pennsylvania, Philadelphia, PA 19104 USA. E-mail: metzler@upenn.edu.

Manuscript received January 28, 2013.

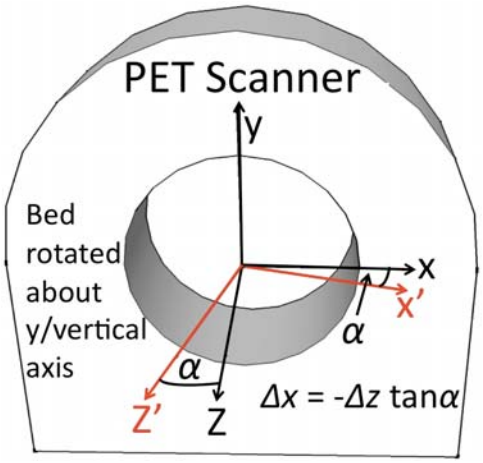


Fig. 1. Right-handed coordinate system describing the PET-scanner geometry. The z axis is aligned with the scanner's axial direction. The bed's in/out direction (z') is in an $x-z$ plane, with angle α between its direction and the z axis; the primed axes (red) are rotated by α about the y (vertical) direction. The value of α is exaggerated for clarity.

II. Methods

A. Super-Sampling with Near-Continuous Bed Motion

Figure 1 shows the right-handed coordinate system used in this paper. The z axis corresponds to the scanner's axial direction. The y axis is vertical, corresponding to the bed's up/down direction. The x axis is in the horizontal direction. The z' axis corresponds to the bed's in/out direction. The x' axis corresponds to the side-to-side direction, which typically is not under motorized control. The x' and z' axes make an angle α with the x and z axes, respectively.

(a) (b) (c) (d) (e)

Fig. 2. (a) A high-resolution image (512x512; 0.125 mm pixels) of a hot-rod phantom (50 mm diameter; rod diameters: 1.2, 1.6, 2.4, 3.2, 4.0, and 4.8 mm) on a warm background (hot:back=5:1). The image is blurred (2.0 mm FWHM Gaussian) and down-sampled to 64x64 (1 mm) pixels in (b)-(e), but with relative shifts in the horizontal and vertical directions (in mm) of (0,0), (-2.5,0), (0,-2.5), and (-2.5,-2.5), respectively. This half-integer bin shifting (2.5 instead of 2) results in averaging different high-res pixels in (a) into lower resolution pixels for (b)-(e).

In a typical PET system, $\alpha = 0$. Consider a point (e.g., a voxel) that does not move relative to the bed. When the bed moves in/out, the voxel's z and z' coordinate values change by the same amount, but the transverse coordinates relative to the scanner (x, y) are unchanged. By introducing a small, non-zero angle for α, we can change the voxel's x position relative to the scanner's coordinate system when the scanner moves in and out. Combining this change in x when z changes during small in/out shifts (i.e., nearly continuous bed motion) with coordinated, small up/down (y) shifts, we can obtain many samples for a voxel at different shifts relative to the scanner; this is exactly what is needed for super-sampling and is equivalent to wobbling by known amounts.

B. Super-Sampling of PET Images

PET reconstructions are digital images, which inherently average over a small area or volume to produce pixel/volume values. Figure 2 shows several examples of the effect of that averaging by shifting the edge of the first pixel between down-samples ((b)-(d)), and adding additional blurring (2 mm). These images can be thought of as being reconstructions with slightly different positioning of the object. We apply a maximum-likelihood estimation-maximization (MLEM)-type algorithm to image-process these four samples – by including the relative shift information – to reconstruct a single composite image.

III. 3D Sampling Uniformity

We evaluated the 3D quality of the sampling for different possible configurations, including bed vertical stepping and different values of α. There are numerous ways to calculate this uniformity. Herein, we used as our metric the normalized standard deviation of the weighted distances of super-sampling points in a unit cell to all points in that cell:

$$\text{Sampling Uniformity} = \frac{\sqrt{\overline{d^2} - \overline{d}^2}}{\overline{d}}, \tag{1}$$

where $d(\vec{r})$ is the weighted distance of super-sampling points (e.g., there are 64 super-sampling points in 8^2 sampling) to point $\vec{r}$:

$$d(\vec{r}) = \sum_s e^{-(\vec{r}-\vec{r_s})^2/(2\sigma^2)} \tag{2}$$

with $\sigma = 0.1$mm; $\vec{r_s}$ is the position of sampling point s in the unit cell and the sum is over all sampling points. We define

$$\overline{d^n} = \frac{\int d^n(\vec{r})dV}{\int dV} \tag{3}$$

as the volume-averaged value over the unit cell. This metric, although somewhat arbitrary, is related to the number of unique samples of a voxel.

A. Reconstructions of 2D Super-Sampled PET Acquisitions

We reconstructed simulated 2D acquisitions of a hot-rod phantom on our whole-body research scanner, La-PET, which has 4 mm $\times$ 4 mm $LaBr_3$ crystals in a 93 cm ring. The simulation assumed perfect detection (i.e., no degradation beyond position uncertainty from the crystal size). The phantom was moved through different sampling positions over a 2D unit cell. We considered four sampling configurations: 1^2, 2^2, 4^2, and 8^2, where the base indicates the number of steps over a unit cell (i.e., a voxel) in each direction and the exponent indicates the number of orthogonal directions sampled during the scan. For example, 4^2 indicates 4 steps, each 1/4 of voxel in both the horizontal and vertical directions, for a total of 16 acquisition positions. The phantom had hot rods with diameters: 1.2, 1.6, 2.4, 3.2, 4.0, and 4.8 mm on a warm background (hot:back = 5:1).

The reconstruction used a standard MLEM algorithm and incorporated the known object shifts. A total of 1000 iterations were run for each sampling case and for each noise level (10^5, 10^6, 10^7, 10^8, and 10^9 total counts).

B. Quantitative Analysis of Reconstructions

Reconstructions were evaluated for quantitative accuracy by employing normalized RMS error (NE) [11] over the image, adding a scaling factor k. This metric is defined for a reconstruction with value v_i in voxel i in comparison to that voxel's true value, r_i:

$$NE = \frac{\sqrt{\overline{(kv - r)^2}}}{\sqrt{\overline{(r - \overline{r})^2}}} \tag{4}$$

where the overline indicates average values over the voxels and k scales the reconstruction to the reference for fair comparison

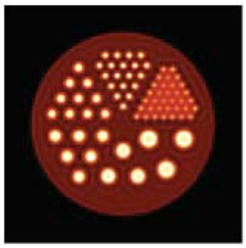


Fig. 3. Super-resolution using post-reconstruction algorithm on images (b)-(e) in Fig. 2.

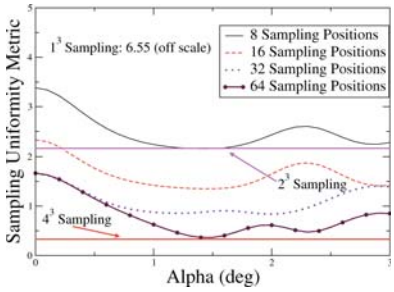


Fig. 4. Sampling Uniformity Metric vs. α for different possible acquisition parameters. These represent samples of the full parameter space, which needs further exploration. Smaller values of this metric indicate better sampling uniformity. This calculation is for one particular metric, described in the text, further exploration of the choice of metric is also required. Note that the curves have more than one minimum, often because of sampling periodicity. A small value for leads to a negligible loss in field of view. For comparison, the horizontal lines indicate the metric values for 2^3 and 4^3 uniform sampling.

between iterations, count level, and sampling techniques. The value of k is that which minimizes NE:

$$k = \frac{\overline{vr}}{\overline{v^2}}. \quad (5)$$

IV. Results

A. Super-Sampling of PET Images

Figure 3 shows the result of post-reconstruction super-resolution using an MLEM algorithm to combine the four samples in Fig. 2((b)-(e)), each on a 64x64 grid, back into a 512x512 grid (as in the high-resolution image (Fig. 2(a)). Please note that the original resolution is not fully restored – due to Gaussian blur and information loss from down-sampling – but the image's improved resolution signifies information recovery.

B. 3D Sampling Uniformity

Figure 4 shows the Sampling Uniformity Metric for several shift sequences. The horizontal reference lines (e.g., 2^3 and 4^3 Sampling) are the theoretical values for grids that are spaced exactly uniformly. The curves vs. α show how the 3D sampling changes as the bed direction changes for different numbers of samples. These curves include the constraint of using only the two motors available on a PET bed. The results show that 3D sampling uniformity comparable to the theoretical limit is achievable using only 2 degrees of freedom.

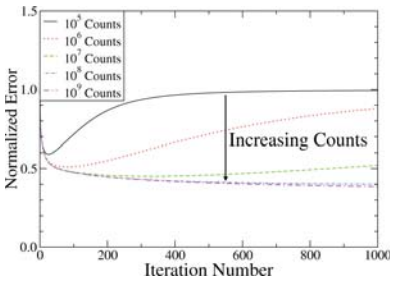


Fig. 5. Normalized RMS error for the reconstructions compared to the digital phantom as a function of iteration number for different count levels. This example is for 1^2 sampling but other samplings show a similar pattern. Note that each count level shows a minimum for normalized error, but the minimum can occur at different iterations.

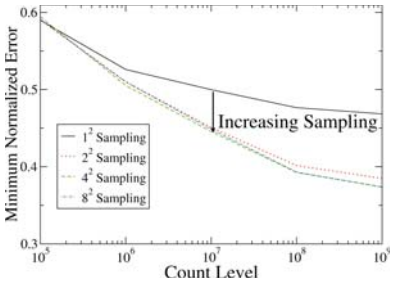


Fig. 6. The minimum normalized RMS error as a function of count rate for different sampling levels

C. Quantitative Analysis of Reconstructions

Figure 5 shows the Normalized Error, defined in (4), for 1^2 sampling for several different count levels, as a function of iteration number. All the curves initially improve as a function of iteration number but then find a minimum error due to increasing noise; higher count levels require more iterations to achieve this minimum error.

Figure 6 shows the value of the minimum error for different sampling configurations and for different count levels.

D. 2D Reconstructions of Super-Sampled PET Acquisitions

Figure 7 shows non-TOF reconstructions of simulated 2D acquisitions on LaPET of a hot-rod phantom in a uniform background. The rows represent different count levels, where the background level in a single 4-mm slice of a clinical scan would often correspond to being between the top and middle rows. The bottom row is a high-count limit, which would be typical for longer scans of single organs (e.g., brain), or for phantom studies.

Columns indicate the amount of super-sampling. The first column has only 1 acquisition. Other columns boost the number of sampling positions in each of the 2 dimensions (e.g., 8^2 sampling is 64 samples on an 8x8 uniform grid). Each reconstruction is shown at the iteration minimizing its normalized RMS error (NE) [11].

V. Discussion

We have taken two reconstruction aproaches. The first is demonstrated in Fig. 3, where images equivalent to four different reconstructions (Fig. 2) with small shifts of the object

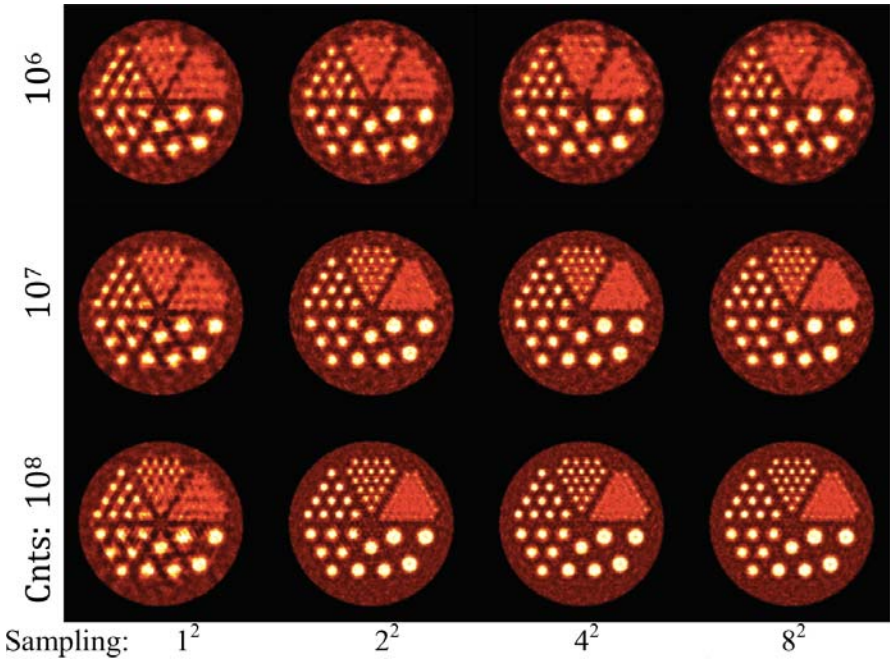


Fig. 7. 2D reconstructions of 50-mm hot-rod phantom (same as in Fig. 2(a)) simulated acquisitions on LaPET. The rows represent low-, medium- and high-count acquisitions, respectively. The columns are for different 2D sampling configurations. 1^2 indicates 1 sample. 8^2 indicates 8 sampling positions in both x and y (64 total). The rods seen in the smallest sector for the 1^2 sampling are artifactual, not occurring in the same position as the phantom. Higher counts levels also allow recovery of finer structures, agreeing with [10].

in between were combined using an MLEM algoritm. The resulting image has much higher resolution indicating that there is greater information content in the set of images than in any individual image.

Figure 3 shows that the method of introducing a small angle ($\sim$1.5°) for the bed direction, combined with vertical movements of the bed, yields sampling uniformity similar to that obtained from exactly regular spacing, which would require an extra degree of freedom in the bed's motion.

Figure 5 shows that the metric we are using to evaluate reconstruction accuracy has a minimum as a function of iteration. Figure 6 compares different sampling densities and count levels at that iteration. For all sampling cases, the error decreases with increased counts, due to lower noise in the data. As the count level increases, the difference in error between 1^2 sampling and the other cases increases. This is because the additional information from super-sampling aides the reconstruction. The advantage seems to saturate quickly with little difference between 2^2 and 4^2 sampling; 4^2 and 8^2 are nearly identical.

Figure 7 shows the reconstructions with different sampling and numbers of counts. At low counts (10^6) there is little difference between the reconstructions, but at higher count levels, the differences become more apparent [10]. Although the 1^2 sampling seems to show rods in the smallest sector, these are located in the wrong place; they are artifactual due to insufficient sampling. On the other hand, the rods in the smallest sector are correctly located, and partially resolved in the columns with higher sampling. There are also advantages seen in the next-to-smallest sector (1.6 mm rods). The difference between 2^2 and higher sampling is not evident, suggesting that the limitation is the detector's spatial resolution. Moreover, sampling is a limitation for the 1^2 (i.e., typical scan) case.

VI. Conclusion

Sampling uniformity similar to complex 3D motion can be achieved by combining a small-angle rotation with axial and vertical bed shifts. The increased uniformity improves reconstruction accuracy and reconstruction resolution.

References

[1] C. Bohm, L. Eriksson, M. Bergstrom, J. Litton, R. Sundman, and M. Singh. "Computer-Assisted Ring-Detector Positron Camera System for Reconstruction Tomography of Brain." *IEEE Trans. Nucl. Sci.*, vol. 25(1), pp. 624–637, 1978.

[2] N.A. Mullani, P.M.M. Te, C.S. Higgins, J.T. Hood, and D.C. Ficke. "Engineering aspects of PETT V." *IEEE Trans. Nucl. Sci.*, vol. 26(2), pp. 2703–2706, 1979.

[3] S.R. Cherry *et al.*. "MicroPET: A high resolution PET scanner for imaging small animals." *IEEE Trans. Nucl. Sci.*, vol. 44(3), pp. 1161–1166, 1997.

[4] A.F. Chatziioannou *et al.*. "Performance evaluation of microPET: A high-resolution lutetium oxyorthosilicate PET scanner for animal imaging." *J Nucl Med.*, vol. 40(7), pp. 1164-1175, 1999.

[5] A.F. Chatziioannou, R.W. Silverman, K. Meadors, T.H. Farquhar, and S.R. Cherry. "Techniques to improve the spatial sampling of MicroPET-a high resolution animal PET tomograph." *IEEE Trans. Nucl. Sci.*, vol. 47(2), pp. 422–427, 2000.

[6] J. Verhaeghe and A.J. Reader. "A PET supersets data framework for exploitation of known motion in image reconstruction." *Med Phys.* vol. 37(9), pp. 4709-4721, 2010.

[7] J.Y. Suk, C.J. Thompson, A. Labuda and A.L. Goertzen. "Improvement of the spatial resolution of the MicroPET R4 scanner by wobbling the bed." *Med. Phy.*, vol. 35(4), pp. 1223-1231, 2008.

[8] K.Y. Jeong, K. Choi, W.H. Nam, and J.B. Ra. "Sinogram-based super-resolution in PET." *Phys Med Biol.* vol. 56(15), pp. 4881-4894, 2011.

[9] C.J. Thompson, S. St James, and N. Tomic. "Under-sampling in PET scanners as a source of image blurring." *Nucl Instrum Meth A.* vol. 545(1-2), pp. 436-445, 2005.

A Dynamic Regularization for Time Activity Curve Estimation in Dynamic Pinhole SPECT

Mahmoud Abdalah[1], Rostyslav Boutchko[2], Debasis Mitra[1], Shi Chen[1], Grant T. Gullberg[2]

Abstract: **In this work we propose a new regularization algorithm to an optimization problem of generating input blood-activity function from dynamic imaging in SPECT. The algorithm exploits masking with anatomical structures of the imaged subject from a reconstruction of consistent projection data of later tomographic acquisitions. We develop a hybrid approach, where we use a b-spline method to obtain the initial estimates of the time-activity curves and then, those initial estimates are fed to the FADS (factor analysis of dynamic structures) approach for further refinement. Furthermore, we also include a new technique of auto-updating of regularization weighting-parameters within the iterations of the algorithm. Finally, the proposed method is tested and the results are shown for SPECT data from a NCAT simulation and a rat study.**

***Index Terms*: Dynamic SPECT, Image Reconstruction, Regularization, Optimization, Regularization parameter selection.**

I. INTRODUCTION

Dynamic SPECT imaging aims to determine the kinetic parameters of wash-in and wash-out of a tracer for different tissue types that are imaged [1]. This needs accurate estimation of blood input functions or time activity curves (TAC). Our work involves finding these functions from inconsistent projection data over the first rotation when the tracer is being injected. However, this is a severely under-determined problem.

For a given vector $\boldsymbol{P}$ of measured data (i.e. sinogram), a system matrix $\boldsymbol{S}$, and the imaged volume/subject $\boldsymbol{V}$, dynamic SPECT problem can be modeled by the following system of equations:

$$\boldsymbol{P_{nm} = S_{nm,k} V_{k,m}} \qquad (\mathbf{1})$$

where $\boldsymbol{m}$ is the projection or frame index, $\boldsymbol{n}$ is the pixel index over all frames, and $\boldsymbol{k}$ is the voxel index of the imaged volume. In our problem of dynamic SPECT-imaging, the projection number $\boldsymbol{m}$ also represents the time points of acquiring the data. Upper case of each index will indicate corresponding maximum number, as a matter of convention.

This work was supported in part by the National Institutes of Health NIH grants R01EB07219 and R01HL50663 and the U.S. Department of Energy Contract No. DEAC02-05CH11231.

Mahmoud Abdalah, D. Mitra, and S. Chen are with the Computer Science Department, Florida Institute of Technology, 171 Cypress Brook cir. 1310, Melbourne, FL, 32901. (Phone: 614-316-6039, Email: mabdalah2009@my.fit.edu). R. Boutchko, and G. Gullberg are with the Dept of Radiotracer Development and Imaging Technologies, Life Science Division of Lawrence Berkeley National Laboratory.

The goal is to estimate $\boldsymbol{m}$ volumes or all $\boldsymbol{V_{k,m}}$ from the measured data $\boldsymbol{P}$ (i.e. sinogram). Clearly, eq. (1) is an underdetermined problem (also, ill-posed) since it demands estimating each volume from a very limited number of projections (one projection, in case a single head is used to acquire each timeframe). In this context, we refer to the volumes $\boldsymbol{V_{k,m}}$ over $\boldsymbol{m}$ as Time Activity Curves (TACs).

To reduce the dimensionality of the problem, time-dependent voxel intensities can be expanded in terms of a small number $\boldsymbol{J}$ of *time basis functions* or *factors*, i.e. at any time point or frame/projection $\boldsymbol{m}$. The intensity in the $\boldsymbol{k^{th}}$ voxel is assumed to be a linear combination of $\boldsymbol{J}$ time basis functions/factors $\boldsymbol{f_j}$:

$$\boldsymbol{V_{k,m} = V_k(t_m) = \sum_{j=1}^{J} C_{k,j} f_{j,m}} \qquad (\mathbf{2})$$

where $\boldsymbol{C}$ is the set of coefficients for the factors $\boldsymbol{f}$.

Plugging eq. (2) into eq. (**1**), we get the factored model for a dynamic sinogram:

$$\boldsymbol{P_{nm} = S_{nm,k} C_{k,j} f_{j,m}} \qquad (\mathbf{3})$$

Typically, the following objective function is minimized to estimate the time basis functions/factors and their coefficients.

$$\boldsymbol{g(c,f) = (SCf - P)^T w^{-1} (SCf - P)} \qquad (\mathbf{4})$$

where $\boldsymbol{T}$ denotes the transpose and $\boldsymbol{w}$ is the weighting diagonal matrix of the estimated noise variance. However, the exact $\boldsymbol{w}$ is unknown since the true projection is unknown in advance. In practice, $\boldsymbol{w}$ is simply set to $\boldsymbol{P}$ or estimated from the measured data [2]. Since the problem is underdetermined, regularization is typically added to the objective function (eq. **4**) and then solved by any minimization technique (e.g. gradient search).

The rest of this paper is organized as follows. In section two, a summary of related work is given. Next, a description of our methods is given in section three. Results of the proposed methods are shown in section four. Finally, a conclusion is drawn in section five.

II. RELATED WORKS

Several methods have been proposed to tackle the problem in equation (**1**), including *b-spline* based methods [3][4], Factor Analysis of Dynamic Structures (*FADS*) methods [1][5], and a *dSPECT* algorithm [6][7].

In the first approach, the time basis functions are set to a chosen set of *b-splines* (typically, cubic splines) that have the capability of representing the temporal behavior of the radiotracer in relevant tissues. Hence, eq. (**4**) is minimized only for the coefficients of those b-splines and then, the voxel values and corresponding TACs can be obtained from eq. (2). Some results of b-spline techniques are shown in [3][4]. Usually, b-spline methods produce smooth TACs even in the presence of a high level of noise. This is because the algorithm is restricted for estimating the coefficients of the chosen b-splines only. However, estimated TACs are highly dependent on the choice of b-splines and not always represent the true TACs precisely [1].

Factor Analysis of Dynamic Structures (FADS) method tries to estimate the time basis functions or factors and their coefficients at the same time by minimizing the objective function in eq. (**4**) for both unknowns. The main issue in this approach is that optimization for two unknowns will not guarantee a unique solution due to the fact that the problem is highly underdetermined [7].

dSPECT method on the other hand, tries to estimate the time activity curve without factoring the time-dependent volumes $\boldsymbol{V_{k,m}}$. The method imposes a temporal regularization constraint that enforces the estimated TACs for each voxel $\boldsymbol{V_{k,m}}$ to obey certain rules. Namely, the activity in a voxel can only follow one of these patterns: it can increase and then decrease once, or only decrease, or only increase to only one peak. However, the method may have reconstructions with poor contrast between tissues due to the fact that the algorithm does not assume knowledge of spatial locations of the dynamic regions corresponding to those constrained curves [1][7].

III. METHODS

Since the problem is highly underdetermined, we first add an anisotropic total variance (ATV) regularization function to problem in eq. (**4**). Second, a hybrid method of b-splines followed by a FADS approach is introduced for estimation of TACs. Third, for automatically adjusting the regularization parameters, an auto-updating method is included.

A. Regularization Functions Used

Our problem of highly underdetermined and noisy data necessitates regularization with constraints in minimizing the objective function in eq. (**4**). We identified three main constraints. First, the coefficients within the region of the same tissue-type (e.g. myocardium) must not have high variation, but the coefficients of different tissues may vary. Second, only one of the coefficients for a voxel that contain only one tissue type (e.g. either contains blood, myocardium, or liver) must have non-zero value and the rest must be zero or close to zero. On the other hand, multi-tissue type voxels, which are the voxels located on the boundaries between tissues, possibly coming from partial volume effect, can be allowed with multiple non-zero values as their coefficients. Third, estimated time basis functions or factors should be as smooth as possible. In our work, non-negativity was enforced by zeroing the respective value.

To enforce the above constraints, three regularization functions were added to the objective function (**4**) as follows:

$$\boldsymbol{g(c,f)} = \frac{(\boldsymbol{Scf-P})^2}{\boldsymbol{w}} + \boldsymbol{\lambda_1\Omega(c)} + \boldsymbol{\lambda_2\Theta(c)} + \boldsymbol{\lambda_3\Phi(f)} \quad \textbf{(5)}$$

where:

1. λ_1,λ_2, and λ_3 are regularization weighting-parameters;
2. $\Omega(c)$ is a penalty function that prevents coefficient mixing. Currently, we use a function similar to the regularization used in [5], which minimizes the dot-product between the coefficient-vectors. The main difference is that $\Omega(c)$ in our method is applied only to the voxels that have undesired coefficient mixing (See the following section for more explanation on how those voxels are identified). If the function is applied to all voxels equally, some voxels will be undesirably forced to go to zero;
3. $\Theta(c)$ is an anisotropic total variation (ATV) function. Again, the function is applied to the voxels that belong to the same tissue. This technique allows the algorithm to smooth the regions within the same tissue type. In this way, all boundaries/edges are preserved and not mixed with other regions;
4. $\Phi(f)$ is a smoothing function that minimizes the first derivative of each time basis function/factor.

As mentioned above, regularization functions are only applied to the voxels that obey certain criteria. This is achieved by creating a non-binary mask that identifies each recognized tissue-type with a unique index on the mask and marks each unrecognized/uncertain region with a different label. In more details, the mask is built from two binary masks with the following steps (we refer to these masks as static and dynamic masks, respectively):

1. **Static Mask**: since the later frames of the acquired sinogram are consistent (tracer has settled by then), and the relevant tissue types appear clearly on them, a static reconstruction is obtained from those frames. Then, a binary mask $\boldsymbol{M^s}$ is created from the reconstructed static image.
2. **Dynamic Mask**: At each iteration of the dynamic reconstruction algorithm (i.e. the minimization of eq. (5)), another binary mask $\boldsymbol{M^d}$ is created from the currently estimated coefficients by thresholding. The size of this mask is equal to the size of the coefficients (i.e. $\boldsymbol{J\times K}$) since there are ***J*** factors and **K** voxels.

Then the intersection of two masks $\boldsymbol{M^s}$ and $\boldsymbol{M^d}$ forms the final non-binary mask $\boldsymbol{M}$ as follows:

$$M_{k,j} = \begin{cases} 0, & M_k^s = M_{k,j}^d = 0 \\ j, & M_k^s = M_{k,j}^d = 1 \\ u, & M_k^s \neq M_{k,j}^d \end{cases} \qquad \textbf{(6)}$$

$$\text{for } i = 1,2,..J, \;\; k = 1,2,..K$$

where $\boldsymbol{u}$ is a constant unique number that labels uncertain regions from the rest of the identified segments.

B. TACs Estimation Method

The b-spline method provides a good initial estimate of the TACs. We cluster these initial curves by the *k*-means

algorithm, each cluster representing a tissue type. Then a representative of each cluster is used as initialization to the FADS approach. The main rationale is that, the FADs approach cannot guarantee the uniqueness of the solution. Therefore, we used the b-spline method to get a good initial starting point that likely puts the result of the FADS algorithm close to the global minimum.

Clustering the time activity curves also provided an automatic way to segment the tissues. The resulted segments were used to initialize the coefficients in the FADS approach. It was important that before clustering, the curves that had very low variation (flat) were considered as background or noise and were excluded from clustering.

C. Regularization Parameter Selection

We used the conjugate gradient (*CG*) algorithm for optimizing the error function (eq. **5**). It was essential to find the optimal values for the regularization weighting (*RW*) parameters. Our experiments show that if these parameters were not selected carefully the estimated TACs would not be accurate. A typical approach is to use the L-curve technique. Unfortunately, the L-curve approach could be used for selecting only one regularization parameter and this method does not guarantee the optimal selection of parameter values [8]. Therefore, a *Balancing Principal* (BP) technique [9] was used to estimate the RW parameter values within each iteration of the CG algorithm. The BP technique tries to keep the balance between fitting the model to data and regularization functions. To show the RW parameter-updating method, we rewrite the objective function as the following:

$$g(c,f) = \chi_{WLS} + \lambda_0\chi_{c-mix} + \lambda_1\chi_{c-TV} + \lambda_2\,\chi_{f-smooth} \qquad (7)$$

where:

- $\chi_{WLS} = \frac{\|P-Scf\|_2}{w}$ Weighted least squares function.
- $\chi_{c-mix} = \left|\vec{C_i}\cdot\vec{C_j}\right|_1$ Anisotropic coefficients' mix function.
- $\chi_{c-TV} = |\mathrm{ATV(C)}|_1$ Anisotropic total variation function.
- $\chi_{f-smooth} = |\nabla f|_1$ Factors' smoothness function.

There are three lambdas (RWs) in the minimized objective function and the updating formula is in the algorithm below (Fig. 1):

1: $\lambda_0,\lambda_1,\lambda_2 \leftarrow \varepsilon;\ \ \gamma_0 \leftarrow 10;$
2: $c^0 \leftarrow 0;$
3: $f^0 \leftarrow \{B-Spline\ functions\};$
// Minimize for coefficients and factors for some iterations
4: $\boldsymbol{C^1} \leftarrow \boldsymbol{arg\ \min_c}\{g(\boldsymbol{c^0},\boldsymbol{f^0})\};$
5: $\boldsymbol{f^1} \leftarrow \boldsymbol{arg\ \min_f}\{g(\boldsymbol{c^1},\boldsymbol{f^0})\};$
6: $\gamma \leftarrow \gamma_0\left(\frac{g(\boldsymbol{C^1},\boldsymbol{f^1})}{0.05\ g(\mathbf{0},\mathbf{0})}\right)^{1/4};$ *// [Reference [9]]*
7: $k \leftarrow 1;$
8: **Repeat**
// Minimize for coefficients and factors
9: $\boldsymbol{C^{[k+1]}} \leftarrow \boldsymbol{arg\ \min_c}\{g(\boldsymbol{c^k},\boldsymbol{f^k})\};$
10: $\boldsymbol{f^{[k+1]}} \leftarrow \boldsymbol{arg\ \min_f}\{(\boldsymbol{c^{k+1}},\boldsymbol{f^k}\};$
// Update the regularization parameters
11: $\lambda_1^{[k+1]} \leftarrow \frac{1}{\gamma}\frac{\chi_{WLS}^{[k]}}{\chi_{c-mix}^{[k]}};$
12: $\lambda_2^{[k+1]} \leftarrow \frac{1}{\gamma}\frac{\chi_{WLS}^{[k]}}{\chi_{c-TV}^{[k]}};$
13: $\lambda_3^{[k+1]} \leftarrow \frac{1}{\gamma}\frac{\chi_{WLS}^{[k]}}{\chi_{f-smooth}^{[k]}};$
14: **Until** *a stopping criterion is satisfied*

Fig. 1. Regularization parameter updating method

We refer to [9] for detailed description of the parameter updating method.

IV EXPERIMENTS AND RESULTS

A. Simulation

First, the proposed technique was tested on a simulated dataset. The dynamic sinogram was created using 80x80x80 NCAT SPECT-phantom (Fig. 3b) with three tissue-types: blood (blue), myocardium (red) and liver (green). TACs used

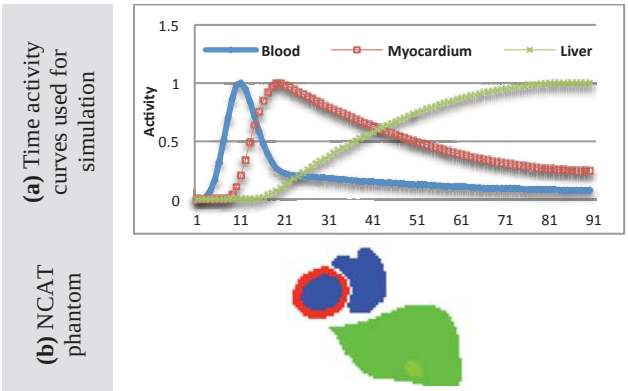


Fig. 2. Simulation dataset.

to mimic the temporal behavior of the tracer in each tissue-type are shown in Fig. 3a. A number of 180 (90 frames/head) frames of size 128x128 bins were generated for 90 seconds (acquisition time is 1 second/frame) using H-shaped dual-head SPECT scanner with a pinhole collimator setting. Several simulation experiments were conducted. First, sets of different initial b-splines were selected to test the hybrid method on each set. The number and shape of each b-spline varied from one set to another. This is to verify how many b-splines are good enough to estimate the initial TACs. Then, for each set/experiment the objective function (eq. 5) was minimized according to the steps mentioned in the previous subsection using the conjugate gradient algorithm.

Fig. 3 shows an example of the results from the algorithm. The first experiment (top row) was performed using a set of two b-splines. Then, the TACs were estimated using the b-spline method (second column) by k=3 clustering. The TACs were then further refined using them as input factors to the FADS algorithm and results are shown in the third column. The algorithm recovers the original TACs used in the simulation (Fig. 2).

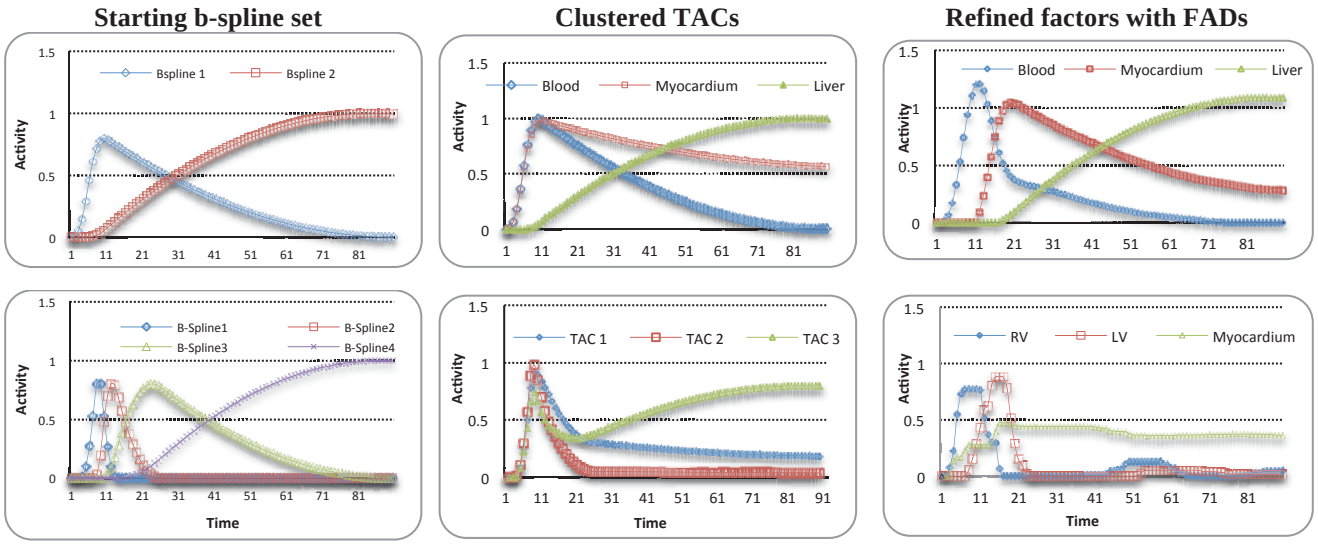


Fig. 3. Results of simulation after 150 iterations (top row) and rat results after 600 iterations (bottom row).

B. Rat Results

The same settings and protocols used to generate the simulated data were also used for imaging a number of SHR and WKY rats. Animals were anesthetized and first injected with 1 mCi of ^{201}Tl and scanned for 12 rotations. Each rotation consisted of 180 angular frames (2 opposing projections acquired every 1 sec). Then, the animals were injected with 5 mCi of ^{123}I-MIBG and imaged for another 60 time frames.

A ROI of blood and myocardium was selected by segmenting the static reconstruction and TACs were estimated from the first rotation using the proposed method. Fig. 3 (bottom row) shows a sample result from a WKY rat. The three factors recovered are from right ventricle, left ventricle and the myocardium. First few seconds' projections were excluded from the input to the FADs algorithm to ensure that noise from the bolus is reduced. Even then, clearly the myocardium still picks up signals from blood. We will need more constraints to be included in the eq. **5** that should model the compartmental kinetics and sparse spatial constraints [**1**].

V. CONCLUSION

A combination of regularization methods is proposed for estimating TACs or input functions in dynamic SPECT imaging. The regularization functions exploit the anatomical structures obtained from the static reconstruction of the later consistent frames, and masks dynamically created from the co-efficients' of factors within the iterations of the optimization algorithm. This helps the algorithm to identify the regions that are in need of regularization. Furthermore, we incorporate a new regularization weighting-parameters selection process to this optimization problem. We are highly encouraged that we could recover tracer-temporal behaviors in blood and tissue regions from such a severely underdetermined problem. Our effort is now directed toward enhancing our set of constraints bringing in more physiological knowledge.

VI. REFERENCES

[1] G.T. Gullberg, B.W. Reutter, A. Sitek, J.S. Maltz, and T.F. Budinger, "Dynamic Single Photon Emission Computed Tomography Basic Principles and Cardiac Applications," *Phys. Med*, Biol. 55, 2010.

[2] Benjamin M.W. Tsui, X. Zhao, E.C. Frey, and G.T. Gullberg. "Comparison Between ML-EM and WLS-CG Algorithms for SPECT Image Reconstruction," *IEEE Transactions on Nuclear Science*, Vol. 38, Issue: 6, pp.1766 - 1772, 1991.

[3] B.W. Reutter, G.T. Gullberg, R. Boutchko, K. Balakrishnan, E.H. Botvinick, and R.H. Huesman, "Fully 4-D dynamic cardiac SPECT image reconstruction using spatiotemporal B-spline voxelization," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 4217–4221, 2007.

[4] B.W. Reutter, G.T. Gullberg, and R.H. Huesman, "Direct Least-Squares Estimation of Spatiotemporal Distributions From Dynamic SPECT Projections Using a Spatial Segmentation and Temporal B-Splines," *IEEE Trans Med Imag*. 19(5) pp. 434–450, 2000.

[5] A Sitek, E. V. R. Di Bella and G. T. Gullberg, " Factor analysis with a priori knowledge—application in dynamic cardiac SPECT," *Phys. Med*, Biol. 45(9), pp. 2619-2638, 2000.

[6] T. Humphries, A. Celler, M. Trummer. "Slow-Rotation Dynamic SPECT with a Temporal Second Derivative Constraint," *Medical Physics*, 38(8), pp. 4489, 2011.

[7] T. Humphries, A. Saad, A. Celler, T. Moller, and M. Trummer. "Segmentation-based regularization of dynamic SPECT reconstructions," *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 2849-2852, 2009.

[8] G. Rodriguez, D. Theis, "An algorithm for estimating the optimal regularization parameter by the L-curve," *Rend. Mat. 25*, pp. 69-84, 2005.

[9] I. Kazufumi, J. Bangti, T. Tomoya, "Multi-Parameter Tikhonov Regularization, " *eprint arXiv*:1102.1173, 2011.

Evaluation of Bouc-Wen Model Corrected Respiratory Motion in Cardiac SPECT

Paul Dasari, Arda Konik, P. Hendrik Pretorius, Mohammed S. Shazeeb, W. Paul Segars, Karen Johnson and Michael A. King

Abstract— **Hysteresis is when the heart follows two different trajectories during inspiration and expiration. In cardiac SPECT imaging, conventional respiratory correction methods based on amplitude binning of a respiratory signal from the abdomen do not take hysteresis into account. This leaves the reconstructed cardiac SPECT images with residual blurring after correcting for respiratory motion. Therefore, prediction of heart motion solely from the abdomen can be suboptimal. We address this problem by implementing the Bouc-Wen (BW) Model of Hysteresis using the signals obtained from external markers placed on both the abdomen and chest. We study the efficacy of the proposed method through analytical simulations of the NURBS-based cardiac-torso (NCAT) phantoms and evaluate the performance of the hysteresis compensated respiratory binning method in comparison with conventional methods. We found that the motion estimates obtained from the BW model were closer to the true respiratory motion of the heart in comparison with the conventional methods for both hysteretic and non-hysteretic cases. The results of our study indicated that our proposed method provides robust motion estimation and could be used for respiratory amplitude binning to reduce image degradation associated with respiratory hysteresis.**

I. Introduction

Respiratory motion causes image artifacts and affects clinical diagnosis in myocardial SPECT images. Conventional respiratory correction methods based on amplitude binning of a respiratory signal (e.g., tracking an external marker or a pneumatic bellow on the abdomen)[1, 2], assumes that the heart follows the same path for both inspiration and expiration. However, studies have demonstrated that certain patients exhibit a pattern in which the heart moves in a different trajectory for the inspiration and expiration phases of the respiratory cycle [3]. Such complex internal behavior suggests that predicting the motion from a single parameter (such as the respiratory signal related to abdominal motion) may not fully describe the motion trajectories of the heart and would likely introduce errors thereby reducing the effectiveness of amplitude binning. Although the respiratory binning methods would help by reducing the blurring due to respiratory motion, the final cardiac SPECT images are left with some residual blurring due to hysteresis thereby degrading lesion detection.

As reported in [4] we investigated the relationship between the anterior-posterior (AP) motions of the external markers on the chest and the abdomen from a Visual Tracking System (VTS). In synchrony with external tracking, the relationship between the superior-inferior (SI) motions of the heart and the diaphragm are obtained using the MR-Navigator technique as internal markers. We concluded from this study that if hysteresis is present internally between the heart and diaphragm similar behavior can be manifested externally in the chest and the abdomen respiratory signals. Such hysteretic behavior can be predicted using an appropriate model for hysteresis based on the acquired external respiratory information. The Bouc-Wen (BW) hysteresis model, desired in mechanical and structural engineering due its ease of numerical implementation, has the ability to represent a wide range of hysteretic loop shapes. Therefore, in this study the BW model of hysteresis is used to describe the hysteretic respiratory motion and hence predict the respiratory motion of the heart devoid of hysteresis.

We used the realistic NCAT [5] human anthropomorphic phantom to evaluate the performance of the proposed hysteresis compensated respiratory binning method in comparison with conventional methods in correcting respiratory motion in cardiac SPECT studies. To model the respiratory motions (including hysteresis) of the heart and other internal organs, internal MRI navigator data from human volunteer studies were used. The known through simulation respiratory motion of the NCAT heart served as the basis of truth for evaluation of the binning methods considered in this study.

II. Methods

The flow chart in Fig. 1 illustrates the steps involved in this study:

Manuscript received April 15, 2013. This work was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under grant R01 EB001457 and a grant from Philips Healthcare. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIBIB or Philips Medical Systems.

Paul Dasari, Arda Konik, P. Hendrik Pretorius, Mohammed S. Shazeeb, Karen Johnson and Michael A. King are with the Department of Radiology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655 (telephone: 508-856-6132, e-mail: paulkreddy@wpi.edu).

W. Paul Segars is with the Department of Radiology *Carl E. Ravin Advanced Imaging Laboratory, Duke University Medical Center, Durham, North Carolina 27705*

Paul Dasari additionally is with the Department of Biomedical Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609.

A. Phantoms

We performed analytical simulations of cardiac perfusion SPECT imaging using the NCAT torso phantom. We simulated hysteretic and non-hysteretic respiratory motion patterns of the liver, heart and spleen in the SI direction. Respiratory motion of the heart is typically smaller in extent than that of the liver and spleen [3, 4, 6]. In addition, a phase difference is observed in the presence of hysteresis [3, 4]. In the default NCAT respiratory motion, organs are restricted to move linearly (i.e., in phase) and a hysteretic pattern is not an available option. In order to accommodate hysteretic pattern, the NCAT phantom was specifically modified to allow the independent motions of heart, liver and spleen. For each of the hysteretic and non-hysteretic respiratory motion patterns, three hundred NCAT phantoms (matrix size: 256×256×256 and voxel size: 0.2332 cm) were generated based on the volunteers' MRI navigator data of the internal organ motions acquired over a period of five minutes. Hence, each NCAT phantom represents an instant of the respiration with realistic organ motions. In addition, average cardiac motion due to beating heart was included based on 16 time frames.

B. Projection Data

We used an analytical projector [7], modeling attenuation (without scatter effect) and distance-dependent collimator blurring (low energy high resolution - LEHR) to obtain 60 projections (radius of rotation: 25 cm) covering 180^o around each NCAT phantom, from 45^o right anterior oblique to 45^o left posterior oblique, representing a cardiac SPECT acquisition in the presence of respiration. Then, 256x256x60 projection data were resized to 128x128x60 and scaled to ~ 7 million total counts to match the Poisson noise level of a ^{99m}Tc-Sestamibi cardiac SPECT study. With this set-up, unlike the variation seen in patient breathing, the 300 NCAT respiratory states were repeated for every projection angle due to the limited number of respiratory samples.

C. Bouc-Wen Model of Hysteresis and Hysteresis Compensation

The respiratory motion information from the chest and the abdomen regions is incorporated in the modified Bouc-Wen model, which is used to model hysteresis in predicting the heart motion.

The modified BW model is expressed as following [8]

$$\dot{w} = \rho\ (\dot{x} + \delta\ x\,|\dot{w}| - \sigma\,|\dot{x}|\,w - \gamma\ w\,|\dot{w}|) \qquad (1)$$

where ρ, δ, σ and γ are parameters that determine the shape and extent of hysteresis in modified Bouc-Wen model, *x(t)* and *w(t)* are the anterior-posterior (AP) displacements of the abdomen and chest markers, respectively.

As previously illustrated in [4], the variation of the respiratory motion between the chest and the abdomen is similar to that of the heart and the liver. We are interested in the variation of respiratory motion between the chest and the abdomen regions, rather than the respiratory rates. Hence, the term *dw/dx*, conveying the variation of respiratory motion between the chest and the abdomen region in (2) is derived by rearranging (1) and dividing it by the respiratory rate *dx/dt*:

$$w_{pred} = \frac{dw}{dx} = \frac{\rho\ (1 - \sigma\ \ w\ \mathrm{sgn}(x)\ )}{(1 - \rho\ \delta\ x\ \mathrm{sgn}(w) + \rho\ \gamma\ w\mathrm{sgn}(w)\ )} \qquad (2)$$

The differential term *dw/dx* represents the predicted data points in the hysteretic loop, which describes the heart motion relative to the abdomen. We then optimized the model parameters (using parametric optimization algorithm, MATLAB) for each respiratory cycle to minimize the error *E* in (3), the difference between the model-predicted heart motion (w_{pred}) and the experimental value of the chest marker.

$$E = \sum_{n=1}^{N} (w_{pred}(n) - w(n))^2 \qquad (3)$$

Using the respiratory motion signal acquired from the external VTS we determined the respiratory motion of the heart based on the BW model fit. Two sets of volunteer respiratory data, one depicting a hysteretic and the other non-hysteretic case, were used for the data fitting in this study.

D. Amplitude Binning and Application of RM Correction

Amplitude binning was performed for four different cases: 1) ideal case which used the actual heart displacements of the NCAT phantom; 2) using the BW model obtained from the abdomen and the chest markers; 3) using just the chest markers; 4) using just the abdomen markers. For each binning case, the projection data was sorted into nine equal bins between the end-inspiration and end-expiration. The binned projection sets were then reconstructed using an OSEM algorithm with attenuation correction, resolution recovery and motion compensation. Attenuation correction was performed using the respiratory blurred attenuation map averaged over 300 NCAT phantoms. For the estimation of motion, the intermediate respiratory state was considered as the reference state while the remaining respiratory states were considered as motion states. The motion estimates were obtained by registering the motion states to the reference state. Finally, using these estimates motion compensated reconstruction was obtained.

The performance of the BW and conventional methods were evaluated using the motion estimates in comparison with the ideal case.

III. Results and Discussion

In Fig. 2 (a), the correlation between the anterior-posterior (AP) motion of the external chest and abdomen markers shows hysteretic loops with different paths for inspiration and expiration. The BW model fit (in red) demonstrates that our proposed model has captured the relationship between inspiration and expiration accurately (top left). In addition, the SI motion of the heart is correlated with the BW signal and the AP motion of the external chest and abdomen markers are depicted as scatter plots (blue). The BW signal shows a better agreement with the heart motion compared to the signal obtained from the abdomen marker. The chest marker also shows a better agreement with the heart motion.

Fig. 2 (b) shows the plots of respiratory motion estimates (for 10 noise-realizations) from the BW signal, the external chest and abdomen markers compared to the actual displacement of the heart. The respiratory motion estimates for the BW signal appear to be closer to the actual heart respiratory motion estimates compared to that of the external chest and abdomen markers. Furthermore, the trend exhibited by the heart motion estimates from the BW signal shows a greater linear tendency compared to that of the chest and abdomen markers. This indicates that BW method better represents the heart motion compared to conventional methods using abdomen or chest markers only.

Fig. 3 (a) shows the correlation between the AP motion of the external chest and abdomen markers for a non-hysteretic case. The BW model fit (in red) has captured the relationship between inspiration and expiration accurately. The SI motion of the heart is correlated with the BW signal and with the AP motion of the external chest and abdomen markers. Although the external chest and abdomen markers correlate well with the heart, the BW signal still shows a better agreement with the heart motion. However, in this particular non-hysteretic case between the abdomen and chest markers, the chest marker shows a poorer correlation with the heart motion, contrary to the observation made in the hysteretic case. This fact demonstrates that the BW signal is more reliable than the chest and abdomen markers in estimating the respiratory motion of the heart.

Further, the respiratory motion estimates (for 10 noise-realizations) from the BW model, the chest and abdomen markers are compared to the heart are shown in Fig. 3 (b). It is noticeable from the plot that the respiratory motion estimates for the BW signal and the external markers show a linear trend. However, on careful observation the estimates from the BW signal are closest to the true heart motion estimates in comparison to that of the external markers, suggesting a strong correlation between the motions of the heart and the BW signal. This demonstrates the potential of the BW model to predict the non-hysteretic cardiac respiratory motion patterns.

The estimations obtained from the BW model are more reliable for both the hysteretic and non-hysteretic respiratory patterns compared to the conventional methods, which rely on just a respiratory signal from abdominal motion. Thus, the BW model provides a robust approach for the amplitude binning of the cardiac SPECT studies.

IV. Conclusion

We presented in this study a model to correct for hysteresis observed in respiratory motion of the heart by incorporating the Bouc-Wen model. The need for this correction was demonstrated with lifelike simulations using the specially modified NCAT phantom. Our preliminary results from these simulation studies have shown that hysteretic loops from the proposed BW model are in better agreement with the "true" respiratory motions of the heart for both the hysteretic and non-hysteretic scenarios. The successful modeling of hysteresis in the hysteretic and non-hysteretic cases demonstrates the capability of the BW model on describing various hysteretic phenomena. However, these results were based on only two sets of volunteer respiratory data. Future work includes the testing of this method with more respiratory datasets and its evaluation of the diagnostic performance on lesion detection by adding lesions of known sizes at known locations to the NCAT heart. Ultimately, the BW method can be extended to respiratory amplitude binning in clinical studies.

References

[1] D. A. Low, M. Nystrom, E. Kalinin, P. Parikh, J. F. Dempsey, J. D. Bradley, *et al.*, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing," *Medical physics (Lancaster),* vol. 30, p. 1254, 2003.

[2] B. W. Reutter, G. J. Klein, K. M. Brennan, and R. H. Huesman, "Acquisition and automated 3-D segmentation of respiratory/cardiac-gated PET transmission images," *IEEE Nuclear Science Symposium conference record (1997),* vol. 2, pp. 1357-1361, 1996.

[3] K. Nehrke, P. Bornert, D. Manke, and J. C. Bock, "Free-breathing cardiac MR imaging: Study of implications of respiratory motion-initial results," *Radiology,* vol. 220, pp. 810-815, Sep 2001.

[4] P. Dasari, K. Johnson, J. Dey, J. M. Mukherjee, Z. Shaokuan, C. Connolly, *et al.*, "MRI investigation of the linkage between respiratory motion of the heart and markers on patient's abdomen and chest: Implications for respiratory amplitude binning list-mode PET and SPECT studies," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, 2011, pp. 3748-3751.

[5] W. P. Segars and B. M. W. Tsui, "Study of the efficacy of respiratory gating in myocardial SPECT using the new 4-D NCAT phantom," *Nuclear Science, IEEE Transactions on,* vol. 49, pp. 675-679, 2002.

[6] K. McLeish, D. L. G. Hill, D. Atkinson, J. M. Blackall, and R. Razavi, "A study of the motion and deformation of the heart due to respiration," *Medical Imaging, IEEE Transactions on,* vol. 21, pp. 1142-1150, 2002.

[7] T. Pan, D. Luo, and M. King, "Design of an efficient 3-D projector and backprojector pair for SPECT," *Proc. Fully-3D Image Reconstruction Radiol. Nucl. Med,* 1995.

[8] J. Hu, K. R. Qin, C. Xiang, and T. H. Lee, "Modeling of Hysteresis in Gene Regulatory Networks," *Bulletin of Mathematical Biology,* vol. 74, pp. 1727-1753, Aug 2012.

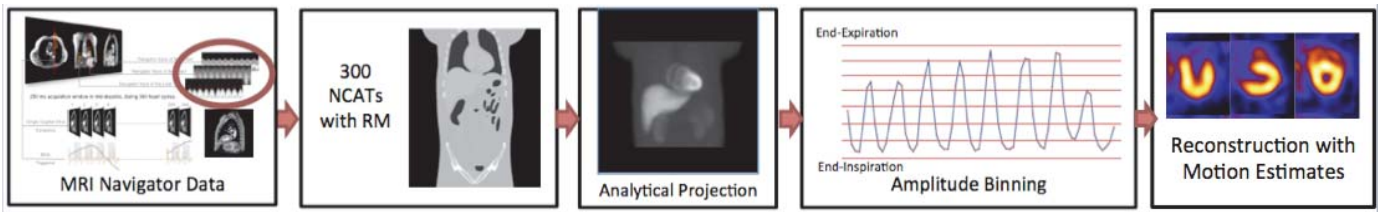


Fig. 1. Flow chart illustrating the simulation of NCAT lifelike respiratory motion with its estimation and correction based on actual respiratory signals. The internal respiratory motions of the organs are captured by the MRI navigator and are then used to create the corresponding organ motion in the NCAT. An analytical projector was used to obtain projections similar to that of the cardiac SPECT studies. Projections were binned based on the corresponding respiratory amplitudes to determine the motion estimates and obtain motion compensated reconstruction.

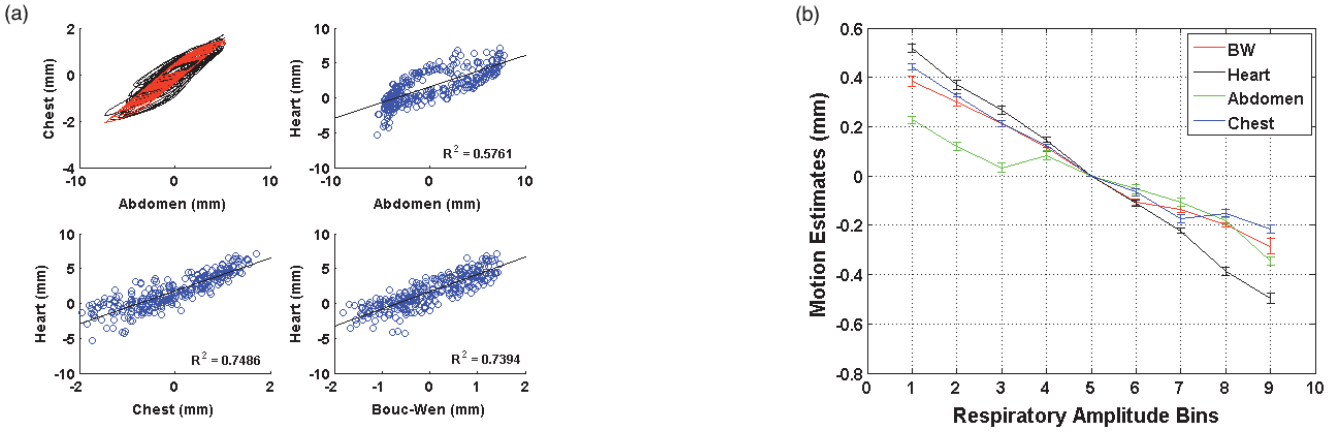


Fig. 2. (a) Scatter plots of the external marker breathing curves (black) and its corresponding BW model fit (red). Also shown are the true heart locations (blue) for a **hysteretic** case versus abdomen, chest, and BW model values. (b) Plot of respiratory motion estimates versus the amplitude bins for respiratory signals obtained from the Bouc-Wen signal (red), the abdomen marker (green) and the chest marker (blue) are compared against the respiratory motion estimates obtained from the actual respiratory motion of the heart (black) which is considered as the ideal binning case.

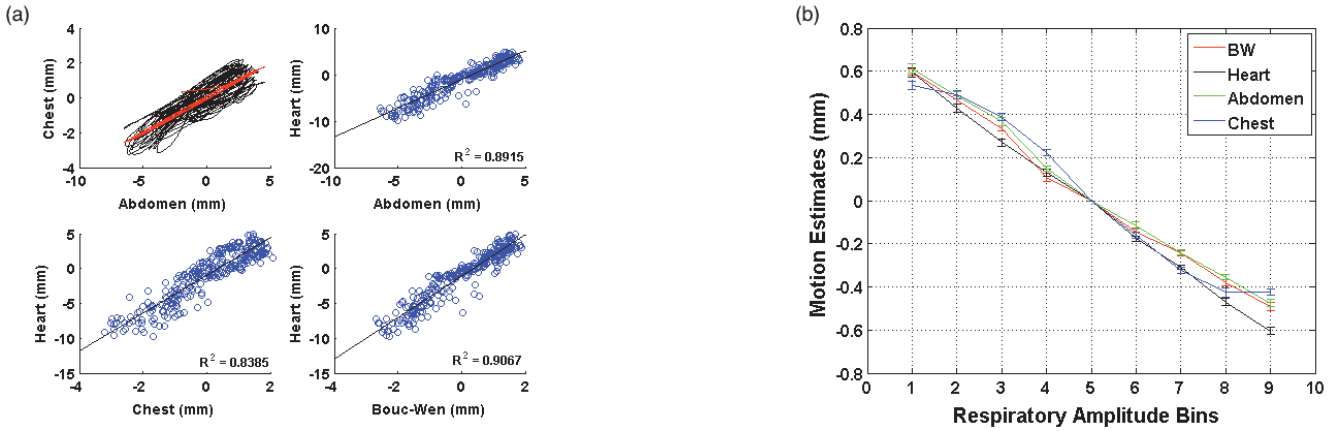


Fig. 3. (a) Scatter plots of the external marker breathing curves (black) and its corresponding BW model fit (red). Also shown are the true heart locations (blue) for **a non-hysteretic** case versus abdomen, chest, and BW model values. (b) Plot of respiratory motion estimates versus the amplitude bins for respiratory signals obtained from the Bouc-Wen signal (red), the abdomen marker (green) and the chest marker (blue) are compared against the respiratory motion estimates obtained from actual respiratory motion of the heart (black) which is considered as the ideal binning case.

Analysis of Ultra-Low Dose CT Acquisition Protocol and Reconstruction Algorithm Combinations for PET Attenuation Correction

Yong Long[1], Lishui Cheng[2], Xue Rui[1], Bruno De Man[1], Adam Alessio[3], Evren Asma[2] and Paul Kinahan[3]
[1]CT Systems and Applications Laboratory, General Electric Global Research Center, Niskayuna, NY, USA
[2]Functional Imaging Laboratory, General Electric Global Research Center, Niskayuna, NY, USA
[3]Department of Radiology, University of Washington, Seattle, WA, USA

***Abstract*—We investigate the tradeoffs between number of views and dose per view in ultra-low dose CT imaging with analytical and statistical reconstruction algorithms for use in PET attenuation correction. We simulated CT and PET imaging scenarios using a 3D NCAT phantom with inserted lesions. CT simulations modeled geometry, Poisson noise and electronic noise while PET simulations modeled geometry, attenuation and Poisson noise. Simulated CT acquisition protocols ranged from 984 views at 1 mA down to 24 views at 41 mA such that the total dose was approximately constant. 20 CT noise realizations were generated for each protocol and images were reconstructed using 3D FDK and OS-SPS using edge-preserving and approximate total variation penalties. Resulting images were then used for PET attenuation correction in OSEM reconstructions of the NCAT phantom. Bias/variance analysis was performed on 10 and 15 mm, 3:1 contrast PET lesions in the lung and liver using 20 PET noise realizations. Simulation results show that the best PET lesion bias/variance tradeoffs with FDK CT reconstructions were achieved with acquisitions of 123 views at 8 mA and that the fidelity of lesion values with FDK reconstruction were highly dependent on CT acquisition technique. On the contrary, OS-SPS CT reconstructions provided PET lesion values that were consistent across acquisition techniques.**

I. Introduction

Combined Positron Emission Tomography (PET)/ Computed Tomography (CT) scanners are in common use for oncology diagnosis and staging [1] where CT images provide anatomical localization and patient attenuation information for PET imaging. When CT images are not intended for diagnostic use and are acquired primarily for PET attenuation correction, the CT radiation dose can be reduced with minimal impact on PET quantitation performance [2], [3]. Overall dose benefits of reducing the CT radiation dose become more pronounced in motion-corrected PET imaging where multiple CT datasets corresponding to different motion gates need to be acquired [4].

In our recent work [4], we investigated selected combinations of dose reduced acquisition methods including spectrum optimization, tube current levels and beam filtration as well as noise reduction techniques such as sinogram smoothing. In this paper we investigate different CT data acquisition and image reconstruction approaches where the total dose delivered to the patient is preserved. We simulated acquisitions ranging from full 984 views at 1 mA down to 24 views at 41 mA.

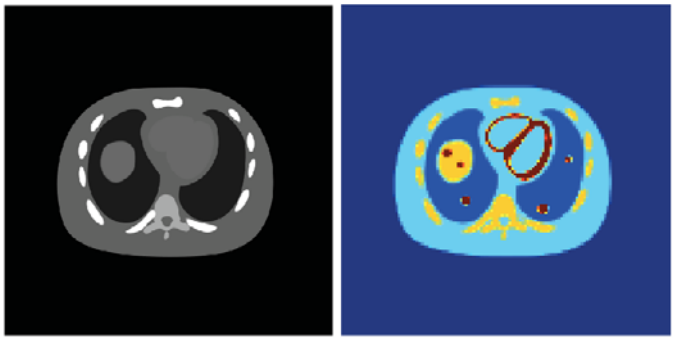


Fig. 1. Central NCAT phantom slices showing the true linear attenuation coefficients at 70 keV (left) and the true activity distribution. All five inserted lesions are visible.

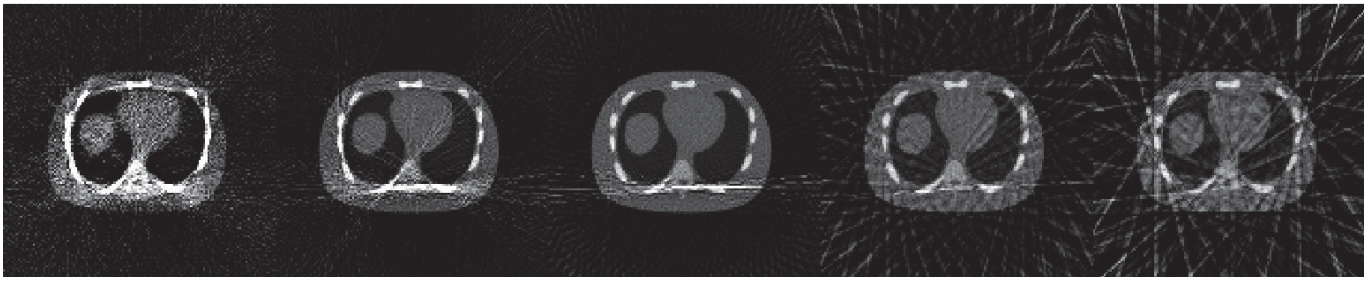
Full view acquisitions do not suffer from missing data but have very noisy data at each view where electronic noise also plays a significant role. Sparser view acquisitions, on the other hand, have higher data quality at each view but need to deal with missing views. Determination of the optimal acquisition protocol provides valuable information on how the dose should be delivered and also depends on the CT reconstruction algorithm.

We used both analytic (FDK: Feldkamp, Davis, Kress) and statistical reconstruction algorithms (OS-SPS: (Ordered Subsets-Separable Paraboloidal Surrogates) for CT reconstructions and statistical reconstructions (OSEM: Ordered Subsets Expectation Maximization) for PET reconstructions. We evaluated acquisition protocols in terms of PET lesion quantitation performance. We accounted for the two random components (attenuation correction factors and photon noise) in PET imaging simulations by generating multiple noise realizations in both CT and PET simulations and analyzing the results statistically across all PET/CT image combinations.

II. Materials and Methods

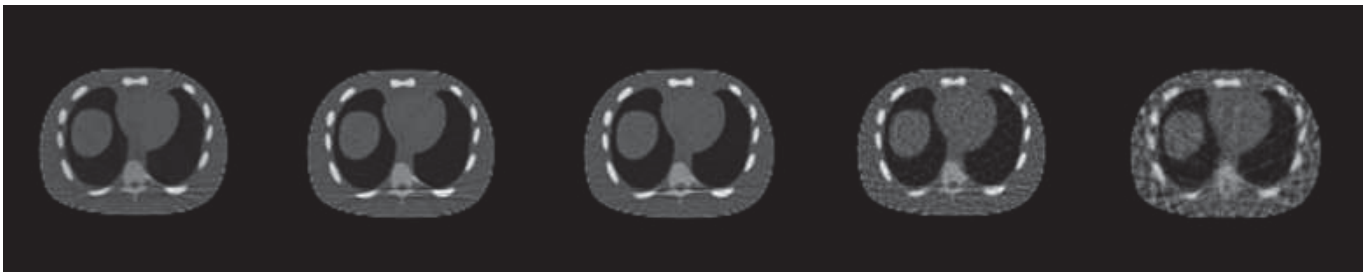
A. CT Simulations

1) Simulation Setup: We simulated a 512×512×72 NCAT phantom with 1 mm isotropic voxels. The 72 mm section of the phantom included lungs and portions of the liver. We



(a) (b) (c) (d) (e)

Fig. 2. Representative central slices reconstructed with FDK for (a): 984 views at 1 mA, (b):328 views at 3 mA, (c):123 views at 8 mA, (d):41 views at 24 mA, (e):24 views at 41 mA. Streaks due to missing data are particularly visible for the 24 and 41 view cases and effects of data noise are dominant in the 1 and 3 mA cases. The linear attenuation coefficient range for all images is 0-0.06 1/mm.



(a) (b) (c) (d) (e)

Fig. 3. Representative central slices reconstructed with OS-SPS using the hyperbola penalty with $\beta = 2^{16}$ and $\delta = 0.2$ for (a): 984 views at 1 mA, (b):328 views at 3 mA, (c):123 views at 8 mA, (d):41 views at 24 mA, (e):24 views at 41 mA. The linear attenuation coefficient range for all images is 0-0.06 1/mm.

used the Computer Assisted Tomography Simulator (CatSIM) [5] to simulate the 3D geometry of the GE VCT scanner with a 70 mm detector, using cone-beam data and a 140 kVp spectrum modeled as approximately monochromatic at 70 keV. The gantry rotation time was 0.35 seconds and we also included electronic noise at 32 electrons per ray. Images were reconstructed directly at the PET resolution of 128×128×11 volumes with 4×4×3.27 mm voxels. Figure 1 shows central slices from the true attenuation and activity distributions used in the simulations and Table I lists the settings that were simulated. We note that at 0.35 mAs these simulations have approximately two orders of magnitude less dose compared to diagnostic CT scans [2], however they result in mean PET quantitation within 10% of the results obtained with full dose CT attenuation correction using 984 views at 500 mA (Figures **??**-**??**).

TABLE I
SIMULATED EQUAL-DOSE CT DATA ACQUISITION SCHEMES

	Number of views	Tube current
Protocol 1	984 views	1 mA
Protocol 2	328 views	3 mA
Protocol 3	123 views	8 mA
Protocol 4	41 views	24 mA
Protocol 5	24 views	41 mA

2) Image Reconstruction: CT images were reconstructed using both analytic and statistical reconstruction techniques. While analytic methods do not model either the compound Poisson or electronic noise encountered at ultra-low dose levels and can not statistically handle sparse views, they nevertheless provide a baseline against which statistical techniques can be measured.

Analytic reconstructions were performed using the FDK algorithm [6] commonly used for small-angle cone-beam reconstructions as in our simulations and statistical reconstructions used the ordered-subsets separable paraboloidal surrogates (OS-SPS) [7] algorithm with the penalized weighted least-squares cost-function (resulting from a second order Taylor Series expansion of the Poisson log-likelihood) and the following hyperbola penalty for which OS-SPS is monotonic:

$$\psi(t) = \delta^2 \left[\sqrt{1 + |t/\delta|^2} - 1 \right] \quad (1)$$

where t denotes the difference between neighboring voxels and δ is a penalty parameter.

Note that for δ values much smaller than typical voxel differences, this penalty approximates the total variation penalty [8]. For larger δ values, it is an edge-preserving penalty that becomes approximately linear at large differences. We used $\delta = 0.2$ and $\delta = 0.0001$ in our reconstructions to evaluate both penalty behaviors.

Neither analytic nor statistical reconstructions used sinogram smoothing techniques, including interpolation for missing data. An intermediate smoothing level (corresponding to $\beta = 2^{16}$) was used for OS-SPS. All reconstructions were directly performed at the PET resolution of 128×128×11 image volumes with 4×4×3.27 mm voxels, which provided significant computational efficiencies for statistical reconstructions.

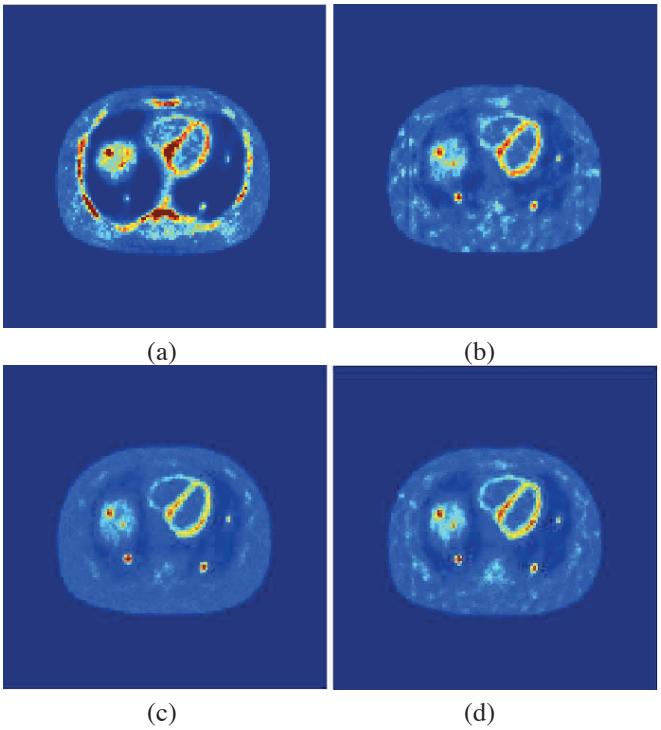


Fig. 4. Representative central PET-image slices from approximately equal dose CT attenuation correction (CTAC): (a) 984 views, 1 mA/view with FDK reconstruction, (b) 24 views, 41 mA/view with FDK reconstruction, (c) 984 views, 1 mA/view with OS-SPS ($\beta = 2^{16}, \delta = 0.2$) reconstruction, and (d) 24 views, 41 mA/view ($\beta = 2^{16}, \delta = 0.2$)with OS-SPS reconstruction. All images have identical colorscales.

Sample FDK and OS-SPS reconstructions for the protocols in Table I are shown in Figures 2 and 3.

Reconstructed images were then converted into linear attenuation coefficient maps at 511 keV by classifying each voxel as air, soft tissue or bone and using the corresponding 70 to 511 keV conversion factors. The resulting attenuation maps were then used to obtain the attenuation correction factors (ACFs) for PET image reconstruction.

B. PET Simulations

1) Simulation Setup: We simulated the fully 3D geometry of the GE Discovery 600 scanner but used a single block ring consisting of 6 axial detectors resulting in 31 direct and oblique planes. Scanner geometry and attenuation effects were modeled. Scatter, randoms, detector blurring and normalization were not modeled as our simulations were focused on the effects of attenuation correction inaccuracies on PET quantitation. 10 and 15 mm diameter lesions were inserted into the NCAT phantom with a contrast of 3:1 with respect to the liver. 5M counts were generated per dataset to simulate a count level of approximately 5 mean counts per non-zero sinogram bin.

2) Image Reconstruction: Each of the 20 different PET noise realizations were reconstructed with each of the 20 511 keV ACFs into 128× 128×11 images with 4×4×3.27 mm voxels using ordered-subsets expectation maximization (OSEM) with 2 iterations, 16 subsets, 4 mm FWHM Gaussian in-plane post-filtering and a 1-6-1 axial smoothing post-filter. As a result, each setting in Table 1 resulted in 400 separate PET images. Figure 4 shows representative PET images reconstructed using attenuation correction factors from CT images reconstructed using FDK and OS-SPS.

3) Quantitation Metrics: Our PET quantitation metrics were the ensemble bias and variances of mean and max activity values (SUV: standardized uptake values) at lesions across PET noise realizations. Note that PET lesion quantitation is affected by two random vectors: random PET data and random ACFs due to randomness in CT data/images. Therefore ensemble mean and variances are calculated via conditional expectation and variance formulas:

$$E[SUV_{mean}] = E_{\mathbf{x}_{CT}}\left[E(SUV_{mean}|\mathbf{x}_{CT})\right] \quad (2)$$

$$\begin{aligned} var(SUV_{mean}) &= E_{\mathbf{x}_{CT}}[var(SUV_{mean}|\mathbf{x}_{CT})] \\ &+ var_{\mathbf{x}_{CT}}[E(SUV_{mean}|\mathbf{x}_{CT})] \end{aligned} \quad (3)$$

where $\mathbf{x}_{CT}$ denotes the CT image and the expectations and variances in the equations are estimated from their respective sample means and standard deviations. SUV_{max} statistics were calculated similarly. These calculations were performed for each of the five CT data acquisition protocols and each of the five inserted lesions.

III. PET Quantitation Results

Figures 5, 6 and 7 show the PET quantitation results for mean activities across all five inserted lesions as a function of CT data acquisition settings for FDK and OS-SPS reconstructions. We see that while FDK reconstructions work better with intermediate acquisition protocols (Protocols 2-4), OS-SPS reconstructions result in approximately equal PET quantitation for all protocols. We observed similar results for max. lesion activity comparisons. It is interesting to note that even though the CT images have visible differences between the protocols, the differences fall within the errorbars. We also note that while the approximate total variation penalty might be able to better handle increasingly fewer views [9], for the protocols under examination, it only provided increased edge preservation which had almost no impact on PET quantitation.

IV. Discussion

We investigated the effects of ultra-low dose CT data acquisition protocols in combination with analytic and statistical CT reconstruction techniques on PET attenuation correction and therefore quantitation. The results of such an analysis depend on a large number of parameters in both CT and PET data acquisition and reconstruction, we fixed all parameters except CT acquisition protocols and reconstruction algorithms to gain insight into better ways of utilizing a fixed, very-low CT dose.

Our results showed that while analytical techniques suffer from either very low dose per view or very few views, statistical techniques are better able to cope with both cases and are more robust to acquisition protocol variations. The

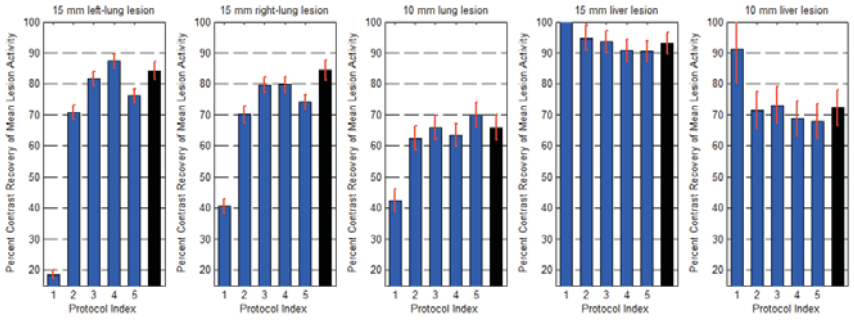


Fig. 5. PET quantitation results for attenuation correction with FDK reconstructed CT images. Black bars correspond to PET quantitation with a full dose CT acquisition with 984 views at 500 mA.

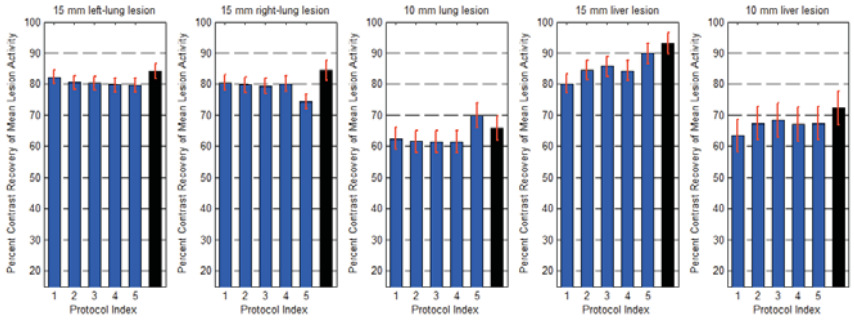


Fig. 6. PET quantitation results for attenuation correction with OSSPS-PWLS reconstructed CT images ($\delta = 0.2, \beta = 2^{16}$). Black bars correspond to PET quantitation with a full dose CT acquisition with 984 views at 500 mA.

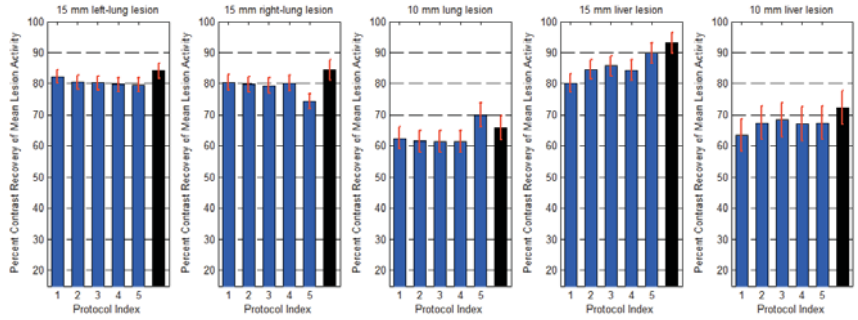


Fig. 7. PET quantitation results for attenuation correction with OSSPS-PWLS reconstructed CT images ($\delta = 0.0001, \beta = 2^{16}$). Black bars correspond to PET quantitation with a full dose CT acquisition with 984 views at 500 mA. Quantitation results are very similar to the $\delta = 0.2$ case indicating that the level of edge-preservation in the CT reconstruction has a minimal effect on PET lesion quantitation.

level of edge-preservation in the CT statistical reconstructions appeared to have very little impact on PET quantitation. We plan to further investigate the effects of other imaging parameters such as TOF-PET reconstruction, smaller and lower contrast lesions to determine if a preference among protocols emerges for statistical reconstructions.

V. Acknowledgements

This work was supported in part by the National Institutes of Health under grant R01 CA160253.

References

[1] D.W. Townsend, J.P.J. Carney, J.T. Yap, N.C. Hall, "PET/CT Today and Tomorrow", Jour. Nuc. Med., (45)1:4S-14S, 2004.

[2] J.G. Colsher, J. Hsieh, J-B. Thibault, A. Lonn, T. Pan, S.J. Lokitz, "Ultra low dose CT for attenuation correction in PET/CT" in Proc. of IEEE Nuc. Sci. Symp. and Med. Img. Conf. pp. 5506-5511, 2008.

[3] T.Xia, A.M. Alessio, P. Kinahan, "Limits of ultra-low dose CT attenuation correction for PET/CT", in Proc. of IEEE Nuc. Sci. Symp. and Med. Img. Conf. pp. 3074-3079, 2009.

[4] T. Xia, A.M. Alessio, B. De Man, R. Manjeshwar, E. Asma and P. Kinahan, "Ultra-low dose CT attenuation correction for PET/CT" Phys. Med. Bio., 57(2):309-328, 2012.

[5] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. Mcolash, P. Sainath, C. Shaughnessy, B. Tower and E. Williams,"CATSIM: a new computer assisted tomography simulation environment", Medical Imaging 2007: Physics of Medical Imaging, Proc. SPIE 6510, 2007.

[6] L.A. Feldkamp, L.C. Davis and J.W. Kess, "Practical cone-beam algorithm", J. Opt. Soc. Amer. A, (1)6:612-619,1984.

[7] H. Erdogan and J.A. Fessler, "Ordered subsets algorithms for transmission tomography, Phys. Med. Biol, 44(11):2835-2851, 1999.

[8] E.Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization", Phys. Med. Biol, 53(17):4777-4807, 2008.

[9] J. Tang,B.E. Nett, G-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms", Phys. Med. Biol, 54(19):5781-5804, 2009.

Direct MAP Estimation of Attenuation Sinogram using TOF PET Data and Anatomical Information

Hao Li, Georges EI Fakhri, and Quanzheng Li

***Abstract*—We propose a one-step Maximum a Posteriori (MAP) method to estimate attenuation sinogram from TOF PET emission data and co-registered anatomical image through TOF data consistency condition and joint entropy based prior. TOF data consistency condition provides the possibility of evaluating gradient of attenuation sinogram from TOF PET emission data. We combine gradient estimation and integration into one step to formulate a data fitting term, and use joint entropy between feature vectors extracted from the anatomical sinogram and attenuation sinogram as a prior term in a Bayersian framework. The feature vectors are defined by scale-space theory to emphasize prominent boundaries, which are likely to present in both sinograms. The resulting problem is solved by conjugate gradient with a backtracking-Armijo line search. Through simulations with PET and MRI images generated from a thorax phantom, we evaluate the performance of this method by comparing it to a two-step estimation method only using TOF data consistency condition. The results demonstrate our one-step MAP method can estimate attenuation sinogram with superior accuracy.**

***Index Terms*—Attenuation estimation, anatomical prior, mutual information, joint entropy, time of flight positron emission tomography, simultaneous PET/MRI.**

I. Introduction

ATTENUATION correction is essential for the quantitation of positron emission tomographic (PET). For a modern PET/computed tomogramphy (CT) scanner, it is usually estimated from an x-ray CT scan. There are several methods proposed to estimate attenuation correction only from emission PET data for variant applications [1] [2]. More interestingly it has been recently shown [3] that attenuation sinogram can be uniquely reconstructed from ideal time-of-flight (TOF) PET emission data through estimating the gradient of attenuation sinogram using TOF PET data consistency condition. A novel analytical two-step method is also developed to compute attenuation correction in [3], however it suffers from unstable estimations of high frequency components and also areas with low photo counts, even without the presence of Poisson noise.

Anatomical information, such as co-registered MR images, could be a possible cure to this problem. Particularly co-registered MR images are readily accessible in simultaneous PET/MRI, which is a multimodality imaging technology that incorporates MR anatomical imaging and PET functional imaging. Multiple attempts [4, 5, 6] have already shown the potential of estimating attenuation image solely using co-registered MR images by segmentation and attenuation coefficient assignment in PET/MRI.

In this paper, we explore the method to improve the TOF PET attenuation-estimation framework in [3] and the possibility to incorporate anatomical information into the framework. We combine gradient estimation and integration into one step to formulate a data fitting term, and use joint entropy between feature vectors extracted from the anatomical sinogram (forward projection of the anatomical image) and attenuation sinogram as a prior term. We evaluate the performance of our method though simulations on a 2D anthropomorphic phantom with joint entropy prior, in comparison with the two-steps method suggested in [3].

II. Methods

A. MAP Estimation Using Anatomical Prior

We denote the attenuated TOF emission data as $m(s,\phi,t)$, where s, ϕ, and t are the transverse distance, projection angle and TOF bin coordinate. The TOF PET data consistency condition [3] is given by the partial differential equation

$$\begin{aligned} &t\,\partial m/\partial s + \partial m/\partial \phi - s\,\partial m/\partial t + \sigma^2\,\partial^2 m/\partial s\partial t \\ &= \left(-mt - \sigma^2\,\partial m/\partial t\right)\partial x/\partial s - m\,\partial x/\partial \phi, \end{aligned} \tag{1}$$

where x denotes the attenuation sinogram and σ is the standard deviation of TOF Gaussian kernel. Slightly modifying the right-hand side of equation (1), we obtain

$$\begin{aligned} &t\,\partial m/\partial s + \partial m/\partial \phi - s\,\partial m/\partial t + \sigma^2\,\partial^2 m/\partial s\partial t \\ &= \left[\left(-mt - \sigma^2\,\partial m/\partial t\right)\partial/\partial s - m\,\partial/\partial \phi\right]x. \end{aligned} \tag{2}$$

Through discretizing this equation and approximating derivatives with finite central difference method, we can obtain a simple linear equation as following

$$B = \left(A_1 D_s + A_2 D_\phi\right)X, \tag{3}$$

where

$$\begin{aligned} B &= \left(TD_s + D_\phi - SD_t + \sigma^2 D_s D_t\right)M, \\ A_1 &= -MT - \sigma^2 D_t M, \\ A_2 &= -M, \end{aligned} \tag{4}$$

and T, S, M, and X denote discretized version of t, s, m, and x, respectively, while D_s, D_ϕ and D_t are the differential operators along corresponding directions. It indicates that attenuation sinogram could be iteratively reconstructed by minimizing a properly defined cost function based the equation (3).

Denoting $R = \left(A_1 D_s + A_2 D_\phi\right)X - B$, the maximum *a posteriori* (MAP) estimate of attenuation sinogram X can be expressed as

$$\hat{X}_{MAP} = \arg\max_{X \geq 0} \frac{p(R|X)p(X)}{p(R)} = \arg\max_{X \geq 0} p(R|X)p(X), \quad (5)$$

where $p(X)$ is the prior probability density function (PDF) of attenuation sinogram X, and $p(R|X)$ is the likelihood function given by noise distribution of R.

Note that noise distributions of A_1 and B are very complicated and difficult to be directly incorporated due to multiple differential operations on Poisson-distributed data M. Therefore as a simplification we approximate the noise of R as Gaussian distributed and use a least-square data fitting term in the log-likelihood function. The resulting cost function is presented as

$$Cost = 1/2 \left\| (A_1 D_s + A_2 D_\phi) X - B \right\|_2^2 - \log p(X). \quad (6)$$

Here we incorporate the information in a co-registered MR image into the prior PDF $p(X)$ through an information theoretic similarity metric $D(X,Y)$ between attenuation sinogram and MR information. Since attenuation reconstruction is primarily being done in sinogram domain, we first forward project MR images into MR sinograms and define $D(X,Y)$ as an information similarity metric between the feature vectors extracted from attenuation sinogram X and MR sinogram Y, respectively. To keep the high-resolution anatomical information of a MR image, this step requires a well-modeled TOF projector program for a TOF PET scanner. The prior PDF $p(X)$ can then be defined as

$$p(X) = \exp(\beta D(X,Y))/N, \quad (7)$$

where N denotes a constant normalization factor and β is the hyper-parameter to control the balance of data fitting and prior information.

Now our cost function can be rewritten as

$$Cost = 1/2 \left\| (A_1 D_s + A_2 D_\phi) X - B \right\|_2^2 - \beta D(X,Y). \quad (8)$$

B. Information Theoretic Similarity Metric

The information theoretic similarity metric $D(X,Y)$ can be defined as mutual information or joint entropy between feature vectors extracted from attenuation sinogram X and anatomical sinogram Y. As joint density priors can usually produce more robust reconstructions than mutual information priors [7], here we define $D(X,Y)$ in terms of joint density between a set of feature vectors as

$$D(X,Y) = -\sum_i H(X_i, Y_i), \quad (9)$$

where X_i and Y_i are the i th feature vectors extracted from X and Y, respectively.

Scale-space method serves as a major basis for extracting feature vectors and utilizing anatomical information as it emphasizes prominent boundaries [8], which are most likely to be correlated in both sinograms. This method generates a set of image features from the original image by blurring it with an increasing filter width, and so gradually wipes out details and underlines prominent boundaries as image scale increases. Since the feature vectors need to be updated in every iteration, we only define three features from both sonograms to simplify the optimization procedure: sinogram itself, moderate Gaussian-blurred sinogram and Laplacian of blurred sinogram. The sum of three similarity metrics from these feature vectors allows us to assign a moderate weight to boundaries information and accelerate the convergence rate at sinogram boundaries.

To compute the joint entropy, the marginal and joint probability density functions are estimated using the Gaussian kernel based Parzen window method [7]. The gradient of joint density can be readily approximated using a fast Fourier-convolution based method suggested in [7]. The non-convexity of joint entropy would not impede convergence of this method as long as the weight of its gradient is not very large compared to the gradient of data fitting term, thus the cost function in (8) can be efficiently minimized using routine gradient based optimization method. Notice the computation of least-square gradient is much more expensive than that of joint entropy gradient and dominates the computational complexity of this method. Since A_1, A_2, and B can be pre-computed from TOF emission data before reconstruction, and both D_s and D_ϕ are sparse matrices, if the emission data size is $N_s N_\phi N_t$, then for each iteration the complexity of gradient estimation is $O(N_s N_\phi N_t)$.

Because the TOF consistency condition based estimation method could only determine the attenuation sinogram to a constant [3], a constant estimation step, such as image segmentation and classification, is required after optimization to correct the constant offset. In this paper, we assume the attenuation values in a small region of the sinogram are given and can be applied to correct the sinogram range. In practice, this constant can be easily computed from a MR image.

III. Numerical Analysis and Results

A. Phantom and simulated data

We evaluate the performance of this method though simulations on a 2D anthropomorphic phantom. Figure 1 shows the true activity (emission) image, attenuation image, and MRI image computed from the same phantom, all of size 256x256 and voxel size 0.11 cm. The attenuation coefficients are 0.032 cm^{-1} for lung, 0.014 cm^{-1} for bone and 0.096 cm^{-1} for

Fig. 1. The true activity image (left), attenuation image (middle), and MRI image (right).

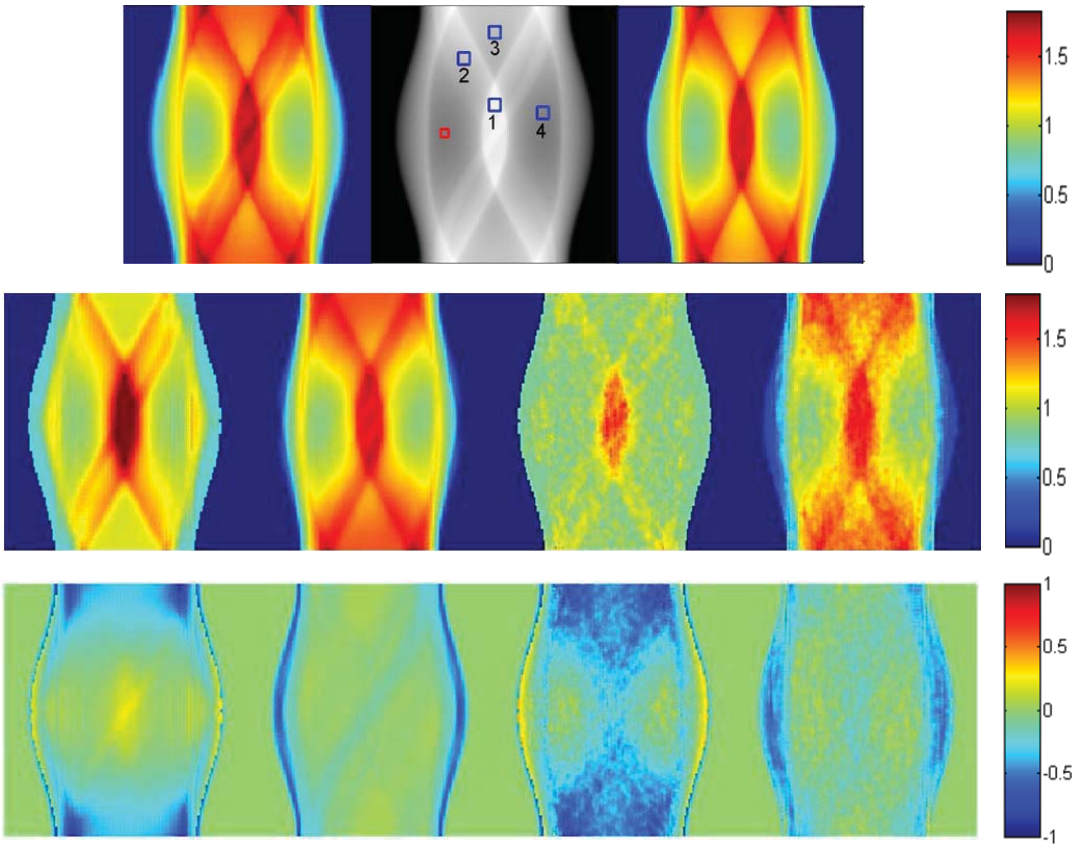


Fig. 2. Attenuation estimation results and bias maps using the two-step method [3] and MAP method with joint entropy prior. Top (L to R): true attenuation sinogram in jet colormap, true attenuation sinogram in gray map with red square of size 5*5 marking the pre-known region and four numbered blue squares of size 7*7 marking regions of interest for bias calculation, MR sinogram; Middle (L to R): estimated attenuation sinogram by the two-step method and MAP method using noise-free TOF data, estimated attenuation sinogram by the two-step method and MAP method using noisy TOF data; Bottom: bias maps of the estimated results in the middle row.

other tissue.

The 2D TOF emission sinogram was generated using a TOFPET forward projector based on Siddon's ray tracing algorithm. The 2D attenuation sinogram and MR sinogram were obtained using the same projector by disabling its TOF function. The radius of scanner ring is 20 cm. The TOF kernel is Gaussian with a spatial FWHM of 7.50 cm (500 ps). The numbers of detectors, rays per angle and TOF bins are 256, 64 and 127, respectively. The TOF data were interleaved into sinogram size 128x128x127 with the pixel size of 1.4° along ϕ, 0.3125 cm along s and 0.23 cm along t, and the attenuation data were interleaved into size 128x128 with the same transverse pixel size. The minimum attenuation percentage is 16.12%. In the attenuated emission data the average and highest photon counts are 0.59 and 2.46, respectively.

Before the estimation of attenuation sinogram, a TOF interval $[t_1, t_2]$ along t was carefully selected to cover the largest object diameter. A sinogram support was acquired by forward projecting the image support, and then further reduced by masking pixels with values below some threshold in any emission sinogram within $[t_1, t_2]$. The noiseless (or noisy) attenuated emission sinogram was smoothed by a Gaussian filter with 2 (or 7) pixels transversal FWHM and 16 pixels TOF FWHM, and the value of σ was then approximated as $\sigma^2 = \sigma_{TOF}^2 + (16\Delta t)^2$, where Δt is the bin interval along t.

B. *Simulaton Results*

The two-steps estimation method in [3] was used as a baseline method. The two derivatives $D_s X$ and $D_\phi X$ were first estimated by performing a least-squares fitting in t within the interval $[t_1, t_2]$. For each derivative, the inaccurate estimations (with very large absolute values) around boundaries as well as their four nearest neighbors in the same s and ϕ were filtered out using the maximum of the absolute values in the central region of the derivative sinogram as a threshold. The attenuation sinogram was then estimated by 2500 iterations of Landweber algorithm, as suggested in [3], and offset-corrected using the given attenuation values in the pre-defined region.

To solve our MAP method with anatomical prior, we used a conjugate gradient (CG) method to minimize the cost function (8). The initial X was estimated using the inverse of trimmed diagonal elements in the Hessian matrix, whose small values below a certain threshold are excluded. The penalty parameter β was 500 for noiseless case and 2500 for noisy case. Four iterations of steepest descent with a vanishing step-size were used to warm up the process, followed by 50 iterations of CG with a backtracking-Armijo line search. At each iteration, three features are extracted from updated sinogram with a gaussian kernel of standard deviation 2.0 pixels and a 3*3 Laplacian

kernel given by $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, which are then used to compute the gradient of joint entropy. The final sinogram is then adjusted with the given constant attenuation value.

Figure 2 shows estimated results and their bias images. It clearly demonstrates the direct MAP estimation using anatomic prior has superior image quality, especially at the thin boundaries. Even for a very noisy data we used in this simulation, we can still estimate a reasonable attenuation sinogram. We also computed the mean relative bias values and their standard deviations in percentage in the whole sinogram as well as four small regions of interest for each estimation result. In the noise-free case, the attenuation sinogram estimated by the MAP method yielded significantly ($p<0.01$) smaller bias (8.2±16.4%) in all regions as compared to the two-step method (-12.2±20.2%). In the presence of Poisson noise, the MAP estimation yielded accurate estimates of the attenuation sinogram (bias of 16.7±12.7%) while the analytical approach departed significantly from the true attenuation distributions (bias of 23.4±25.9%), $p<0.01$. Bias values and standard deviations in four regions of interest are listed in table 1.

The major problem of both methods is incorrect estimation of left and right boundaries of the sinogram. Assigning a larger weight to anatomical prior can reduce this error, but the convergence could be sacrificed due to the non-convexity of joint entropy. A spatially varying weight may be suitable to this situation if necessary knowledge about boundary locations is given. Another problem is that certain kinds of areas in attenuation sinogram, such as the spine-related sine wave from left bottom to right top in the true attenuation sinogram in figure 2, are missing in MR sinogram, due to lower contrast of bone-tissue boundaries in regular MR imaging than in CT scan. In such cases enforcing the maximization of similarity could not provide much help to the reconstruction.

TABLE I
Bias and Standard deviation in ROIs

Method	ROI 1	ROI 2	ROI 3	ROI 4
2-step/noise-free	6.81± 0.01%	-11.36± 0.00%	-20.51± 0.00%	0.03± 0.05%
MAP/noise-free	-1.66± 0.01%	0.60± 0.00%	3.26± 0.00%	-1.11± 0.00%
2-step/noisy	-19.73± 0.13%	-29.41± 0.05%	-32.14± 0.14%	-2.15± 0.14%
MAP/noisy	-7.56± 0.11%	-2.17± 0.04%	-4.23± 0.14%	-0.18± 0.11%

IV. CONCLUTION AND DISCUSSION

The primary goal of this paper is to present a MAP estimation method with anatomical prior based on the newly developed TOF PET attenuation estimation method. We have validated the method though idealized simulations on a simple thorax phantom with comparison to the original two-step method that does not utilize anatomical information. We have shown that in both noise-free and noisy cases, the MAP method achieved superior estimation accuracy and precision of the attenuation sonogram. While a prototype TOF PET/MRI with <400 ps timing resolution is reported in 2012, we believe the estimate results using noisy sinogram can be significantly improved with better TOF resolution, and therefore our method could be a practical approach for the attenuation correction of future clinical PET/MR scanner with state-of-art TOF capability.

The main limitations of this method are that non-convexity of anatomical prior deteriorates the convergence of optimization algorithm if largely weighted and modality mismatch between MR and attenuation images may cause estimation artifact. Possible solutions, such as spatially varying weight, alternative information theoretic metric and attenuation dedicated MR sequence, will be investigated in the future work. Furthermore, the validation in this summary is quite preliminary, we didn't yet evaluate the reconstruction of attenuation and emission images using estimated attenuation sinogram; we will extend our method to 3D and thoroughly validate it using both 3D simulations and clinical data in the future.

REFERENCES

[1] J. Nuyts *et al.*, "Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sonograms," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 393-403, 1999.

[2] A. Rezaei *et al.*, "Simultaneous reconstruction of activity and attenuation in Time-of-Flight PET," *in Conf. Rec: IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, 2011, pp. 2375-2382.

[3] M. Defrise, A. Rezaei, and J. Nuyts, "Time-of-flight PET data determine the attenuation sinogram up to a constant," *Phys. Med. Biol.*, vol. 57, no. 4, pp. 885, 2012.

[4] M. Hofmann *et al.*, "MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration," *J. Nucl. Med.*, vol. 49, no. 11, pp. 1875-1883, 2008.

[5] V. Keereman *et al.*, "MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences," *J. Nucl. Med.*, vol. 51, no. 5, pp. 812-818, 2010.

[6] A. Martinez-Möller *et al.*, "Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data," *J. Nucl. Med.*, vol. 50, no. 4, pp. 520-526, 2009.

[7] S. Somayajula *et al.*, "PET image reconstruction using information theoretic anatomical priors," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 537-549, 2011.

[8] T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales," *J. Appl. Stat.*, vol. 21, no. 1-2, pp. 225-270, 1994.

Soft Classification with Gaussian Mixture Model for Clinical Dual-Energy CT Reconstructions

Ruoqiao Zhang, *Student Member, IEEE,* Jean-Baptiste Thibault, *Member, IEEE,*
Charles A. Bouman, *Fellow, IEEE,* and Ken D. Sauer, *Member, IEEE*

***Abstract*—We study the distribution of the clinical dual-energy CT (DECT) reconstructions by applying a soft classification method with Gaussian mixture model. With a pre-described subclass number, this method estimates the parameters of each subclass and performs classification based on the posterior probability. Our study on the clinical data shows that the classification result relates closely to the actual material composition in the human body, with each material represented by a particular cluster. Also, the study shows that the edges in the DECT images follow a Gaussian mixture distribution, where each subclass has a distinguishable covariance or direction that represents a particular type of edges. Potential usage of this soft classification method includes MRF prior design and accurate material separation.**

***Index Terms*—Computed tomography (CT), dual energy, statistical method, Gaussian mixture, material separation, material classification, Markov random field (MRF) design.**

I. Introduction

DUAL-ENERGY CT (DECT) scanners, which collect X-ray projections with two distinct spectra, are of great interest in applications such as disease diagnosis [1] and security inspection [2]. A DECT reconstruction typically produces cross-sections corresponding to the equivalent densities of two basis materials, where the linear combination of the two uniquely determines the energy dependent attenuation [3]. Typical reconstruction approaches include filtered back-projection (FBP) methods and statistical iterative methods. The statistical approaches allow an accurate model for imaging system and detector noise, which consequently reduce the noise and improve the resolution of the images as compared to FBP.

As for statistical iterative reconstruction approaches, it is critical to build an accurate prior model to represent the image characteristics. This requires knowledge of the distribution of the reconstructed quantities. The Markov random field (MRF) has been applied widely in iterative CT reconstruction as a prior model during the recent past [4]–[6]. Conventional MRF priors for single-energy CT only depend on local pixel differences. However, the distribution of the pixel differences in clinical DECT reconstructions remains unclear. Applying the MRF prior to each basis material components separately, as stated in [7], ignores the correlation between different components. Thus, it is necessary to learn the distribution of the pixel differences in clinical DECT reconstructions. Moreover, since the DECT reconstructions may potentially subject to contamination between different material components, it is also beneficial to enforce material separation during the reconstruction. One possible approach is to introduce material density information in the MRF prior, which is neglected in conventional MRF model. This also requires knowledge of the distribution of clinical DECT reconstructions.

In this paper, we study the distribution of the clinical DECT reconstructions by using a soft classification method with Gaussian mixture model. By using this method, we model the distribution of the reconstructions as a mixture of several distinct subclasses, each of which follows a multivariate Gaussian distribution. Given the total number of the subclasses, we estimate the mean, covariance, and prior probability of each subclass. Based on the estimated parameters, we then classify each data point based on the posterior probability. We use the expectation-maximization (EM) algorithm to solve this problem. The classification result on the clinical data reveals the distribution of neighboring pixel differences and pixel densities of the DECT reconstructions.

This soft classification method can also be used to segment the DECT images. DECT has the potential to determine the materials in the scanned object. Previous classification methods generally threshold the reconstructed values to differentiate the materials [8], [9]. In particular, Zamyatin *et al.* [9] applied a Gaussian-based approach after thresholding to simply determine the boundary of the clusters produced by thresholding. These methods substantially depend on predefined thresholds and may not be robust when two distinct clusters overlap extensively. In contrast, the classification method we use in this paper is fundamentally different from what Zamyatin used in [9], since in this soft classification method the prior distribution of each Gaussian is not uniform and the parameters remain unknown before the algorithm starts. The clinical result of the soft classification method shows a desirable segmentation on the DECT images, especially for highly overlapping soft tissues.

II. Methodology

A. DECT reconstruction

We use the joint dual-energy model-based iterative reconstruction (JDE-MBIR) approach [7] to reconstruct basis material densities. The JDE-MBIR method incorporates a

R. Zhang and C. A. Bouman are with the School of Electrical and Computer Engineering, Purdue University, West lafayette, IN 47907-0501 USA (email: zhang393@purdue.edu; bouman@ecn.purdue.edu).

J.-B. Thibault is with GE Healthcare Technologies, Waukesha, WI 53188 USA (email: jean-baptiste.thibault@med.ge.com).

K. D. Sauer is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556-5637 USA (email: sauer@nd.edu).

The author would like to acknowledge GE Healthcare for supporting this work.

quadratic approximation to the polychromatic log-likelihood with an accurate noise model that fully accounts for the statistical dependency in the decomposed sinograms. This approach has been demonstrated to reduce noise and improve spatial resolution as compared to the filtered back-projection (FBP) and other decomposition-based statistical methods that employ decoupled likelihood model [7].

B. Soft classification with Gaussian mixture model

Let $Y = \{Y_n\}_{n=1}^N$ be a sequence of N multivariate random vectors of dimension M. Each Y_n is modeled by the same Gaussian mixture distribution with K subclasses. Each subclass k is specified by the parameters $\theta_k = (\pi_k, \mu_k, R_k)$, defined as the prior probability, mean, and covariance for subclass k, respectively. Furthermore, let X_n be a random variable that determines the subclass label for Y_n. Then, the conditional probability of Y_n given $X_n = k$ and parameter θ is given by

$$p(y_n|k,\theta) = \frac{1}{(2\pi)^{\frac{M}{2}}}|R_k|^{-\frac{1}{2}} \exp\left\{-\frac{(y_n-\mu_k)^t R_k^{-1}(y_n-\mu_k)}{2}\right\}. \tag{1}$$

Then the conditional probability of Y_n given θ is,

$$p(y_n|\theta) = \sum_{k=1}^{K} p(y_n|k,\theta)\pi_k. \tag{2}$$

The log-likelihood of the entire sequence, Y, is then given by,

$$\log p(y|K,\theta) = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} p(y_n|k,\theta)\pi_k\right). \tag{3}$$

In this paper, we empirically fix the number of subclasses, K. Then the unknown parameter, θ, can be computed as the maximum-likelihood (ML) estimate given by

$$\hat{\theta} = \arg\max_{\theta} \ \log p(y|K,\theta). \tag{4}$$

Due to the unknown state of the labels, $\{X_n\}_{n=1}^N$, we use the EM algorithm to solve this problem. The EM algorithm works by first estimating $\{X_n\}_{n=1}^N$ by clustering the data, $\{Y_n\}_{n=1}^N$, according to the current subclass parameters, θ. Then it reestimates θ based on this clustering result. This leads to an iterative procedure that alternates between classification and parameter estimation. At the i^{th} iteration, the probability that y_n belongs to subclass k can be computed by using Bayes rule,

$$p(k|y_n,\theta^{(i)}) = \frac{p(y_n|k,\theta^{(i)})\pi_k}{\sum_{j=1}^{K} p(y_n|j,\theta^{(i)})\pi_j}. \tag{5}$$

This gives the classification at the i^{th} iteration. This is a "soft" classification since the membership of y_n to each subclass is represented by a probability. Then based on this classification result, we can update the parameters by maximizing the expectation function,

$$\theta^{(i+1)} = \arg\max_{\theta} E[\log p(y, X|\theta)|y, \theta^{(i)}]. \tag{6}$$

A substitution function approach is used to simplify the computation [10]. The updated parameters for each iteration are given by

$$\bar{N}_k = \sum_{n=1}^{N} p(k|y_n,\theta^{(i)}), \tag{7}$$

$$\pi_k^{(i+1)} = \frac{\bar{N}_k}{N}, \tag{8}$$

$$\mu_k^{(i+1)} = \frac{1}{\bar{N}_k}\sum_{n=1}^{N} y_n p(k|y_n,\theta^{(i)}), \tag{9}$$

$$R_k^{(i+1)} = \frac{1}{\bar{N}_k}\sum_{n=1}^{N}\left(y_n-\mu_k^{(i+1)}\right)\left(y_n-\mu_k^{(i+1)}\right)^T p(k|y_n,\theta^{(i)}). \tag{10}$$

The final classification is computed by

$$\hat{X}_n = \arg\max_{k\in K} \ p(k|y_n,\hat{\theta}). \tag{11}$$

The initial condition can be chosen in the same manner as stated in [10],

$$\pi_k^{(1)} = \frac{1}{K} \tag{12}$$

$$\mu_k^{(1)} = y_l, \text{ where } l = \left\lfloor \frac{(k-1)(N-1)}{(K-1)} \right\rfloor \tag{13}$$

$$R_k^{(1)} = \frac{1}{N}\sum_{n=1}^{N} y_n y_n^t \tag{14}$$

where $\lfloor\cdot\rfloor$ takes the greatest smaller integer. Notice that the initial condition does not require any knowledge of each subclass.

C. Data Formulation

We proposed two different formulations of the reconstructed values to study the distribution of a single pixel and its neighbors.

1) Material distances within neighboring pixels: We first study the distribution of the material distance that is measured as the difference within the same material between neighboring pixels. More precisely, let m_1 and m_2 be the reconstructed water and iodine density images, respectively. Furthermore, let s and r be the locations of two pixels in the image. Then we formulate the 2D distance vector as

$$y_{\{s,r\}} = (m_{1,s} - m_{1,r}, m_{2,s} - m_{2,r}).$$

We then formulate a sequence y as a collection of all such 2D vectors over the entire 8-neighborhood system, $\mathcal{C}$. The soft classification is performed on y, with the subclass number, K, empirically fixed to 7.

2) Material densities within neighboring pixels: We formulate a 4D vector that includes the material density information of a neighboring pixel pair. More precisely, for a neighboring pair $\{s,r\}$, the data vector is formulated as

$$y'_{\{s,r\}} = (m_{1,s}, m_{2,s}, m_{1,r}, m_{2,r}).$$

The soft classification is performed on y', which is the sequence of all such 4D vectors over $\mathcal{C}$, with K empirically fixed to 10.

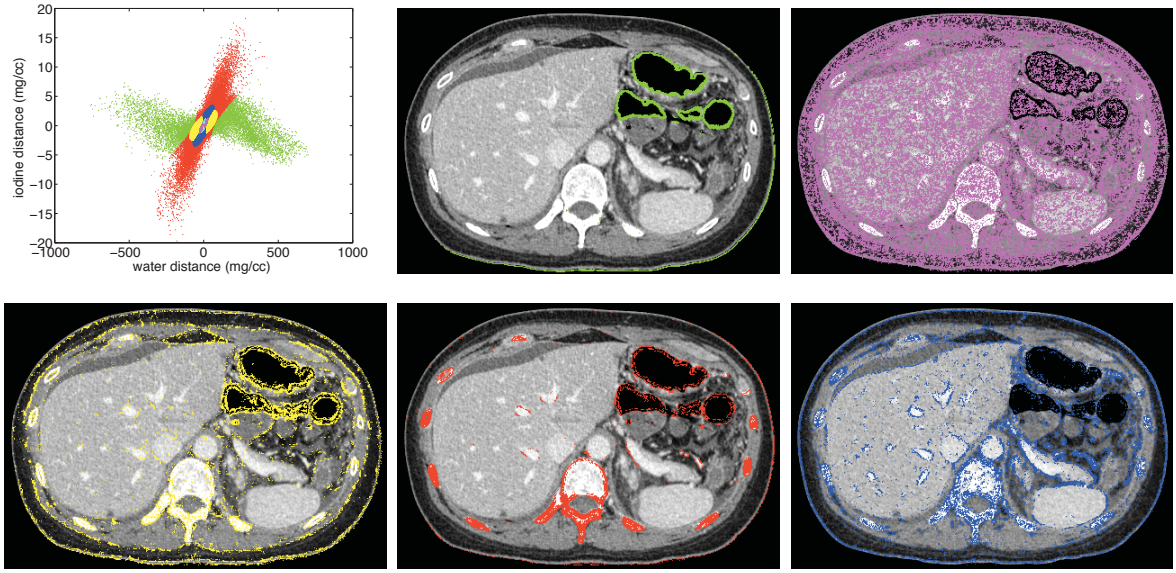


Fig. 1. Classification of the 2D distance vector for FBP images. The scatter-plot shows the classification result, with each color specifying a particular cluster. Then five out of seven subclasses are shown individually on the image, with each represent a particular type of edges. The other two subclasses basically represent DC components and are generally not of great interest.

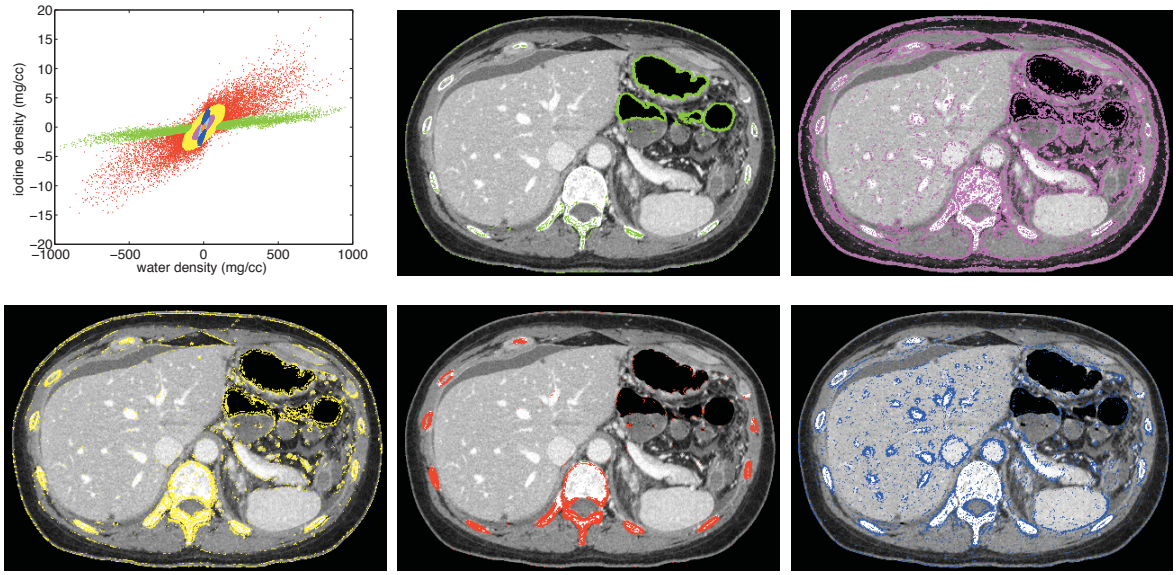


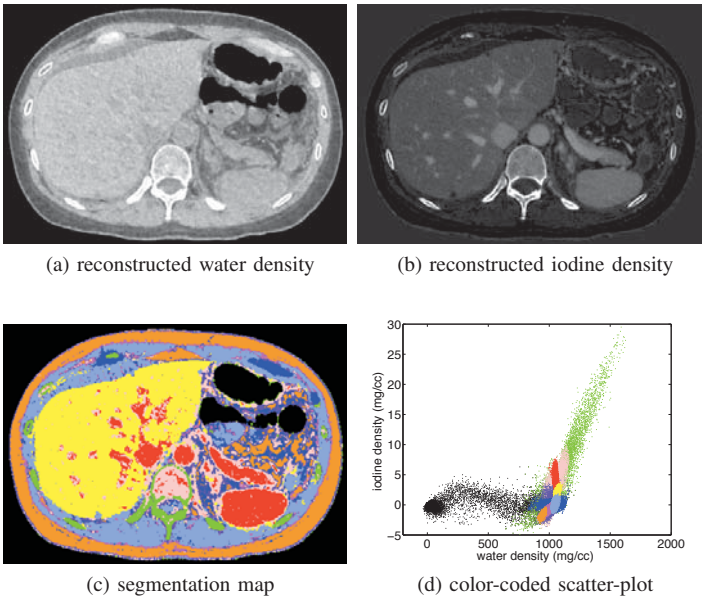
Fig. 2. Classification of the 2D distance vector for JDE-MBIR images. The scatter-plot shows the classification result, with each color specifying a particular cluster. Then five out of seven subclasses are shown individually on the image, with each represent a particular type of edges. The other two subclasses basically represent DC components and are generally not of great interest.

III. Results

We applied the soft classification method on DECT clinical reconstructions. Raw data were acquired on a Discovery CT750 HD scanner (GE Healthcare, WI, USA) in dual-energy fast kVp switching acquisition mode, with tube voltage alternating between 80 kVp and 140 kVp in 540 mAs. We use two methods to reconstruct the water- and iodine-equivalent densities, the FBP method with a standard reconstruction filter kernel and the JDE-MBIR method. We then experimented with the same slice reconstructed by different methods, where the data y and y' were formulated in the manner described in Sec. II-C. We applied the soft classification on y and y' separately.

Fig. 1 and 2 show the classification results on the 2D distance vectors for FBP and JDE-MBIR, respectively. The color-code remains the same for the scatter-plot and the images and the same for FBP and JDE-MBIR cases as well. The results show that the clusters correspond to different types of edges in the images. The clusters have zero mean but different covariances or directions from each other. This suggests an MRF prior that models the distribution of each edge cluster such that different edges can be treated in different ways based on their covariances.

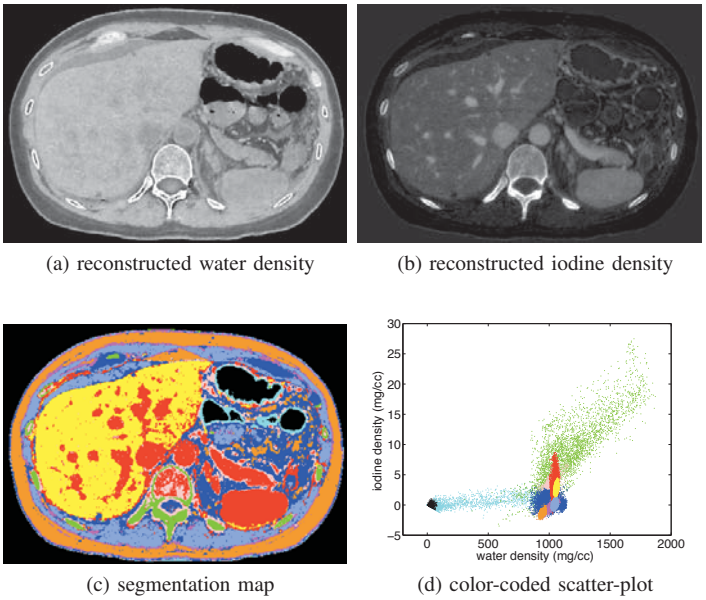
It is also shown in the scatter-plots in Fig. 1 and 2 that



(a) reconstructed water density (b) reconstructed iodine density

(c) segmentation map (d) color-coded scatter-plot

Fig. 3. Classification of the 4D density vector for FBP images. Upper row shows the material density images used in the classification. The classification result is shown on the segmentation map on lower left, with each color specifying a particular cluster. The classification is also shown in the scatter-plot on lower right with the same color code.



(a) reconstructed water density (b) reconstructed iodine density

(c) segmentation map (d) color-coded scatter-plot

Fig. 4. Classification of the 4D density vector for JDE-MBIR images. Upper row shows the material density images used in the classification. The classification result is shown on the segmentation map on lower left, with each color specifying a particular cluster. The classification is also shown in the scatter-plot on lower right with the same color code.

JDE-MBIR changes the distribution of edges as compared to FBP, where the edges between different material components are positively correlated in JDE-MBIR images. The classification result on the JDE-MBIR images captures and separates different type of edges better than that on the FBP images. As shown in Fig. 2, the high contrast edges (bone-tissue edges (red) and air-tissue edges (green)), and the soft tissue edges (magenta and blue), have been well classified. This is because the JDE-MBIR method produces shaper edges and smoother texture than the FBP method [7].

The classification results of the 4D density vectors are shown in Fig. 3 and 4 for FBP and JDE-MBIR, respectively. Both results show that the clusters relate closely to different compositions of human body, such as fat, muscle, bone, blood, and air. As shown in the images, the soft classification method performs well in differentiating highly overlapping tissues. Moreover, the 4D classification also produces clusters that reflect the edges. For example, the cyan cluster in JDE-MBIR result stands for air-tissue edges. These results indicate the possibility of designing an MRF prior that models both the density distribution and the edge distribution to improve material separation and edge performance simultaneously.

IV. Conclusion

We have presented a study on the distribution of the clinical DECT reconstructions by using a soft classification method with Gaussian mixture model. The soft classification method estimates the parameters of each subclass and performs classification based on the posterior probability. Clinical results have shown that the classification results relate closely to different types of edges in DECT images and different body compositions. Future investigation includes accurate material separation and correlation-based MRF prior design.

References

[1] T. R. Johnson, B. Krauss, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard, C. Fink, S. Weckbach, M. Lenhard, B. Schmidt, T. Flohr, M. F. Reiser, and C. R. Becker, "Material differentiation by dual energy CT: initial experience," *European J. Radiology*, vol. 17, no. 6, pp. 1510–1517, Jun. 2007.

[2] S. Singh and M. Singh, "Explosives detection systems (EDS) for aviation security," *Signal Process.*, vol. 83, no. 1, pp. 31–55, Jan. 2003.

[3] L. A. Lehmann, R. E. Alvarez, A. Macovski, W. R. Brody, N. J. Pelc, S. J. Riederer, and A. L. Hall, "Generalized image combinations in dual KVP digital radiography," *Med. Phys.*, vol. 8, no. 5, pp. 659–667, Sept./Oct. 1981.

[4] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[5] ——, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 480–492, Mar. 1996.

[6] J.-B. Thibault, K. D. Sauer, J. Hsieh, and C. A. Bouman, "A three-dimensional statistical approach to improve image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, Nov. 2007.

[7] R. Zhang, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A model-based iterative algorithm for dual-energy X-ray CT reconstruction," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 439–443.

[8] N. J. Pelc, "Dual Energy: technical curiosity or potential clinical tool," in *RSNA*, 2007.

[9] A. A. Zamyatin, A. Natarajan, and Y. Zou, "Advanced material separation technique based on dual energy CT scanning," in *Proc. SPIE 7258, Medical Imaging 2009: Phys. Med. Imag.*, vol. 7258, 725844, 2009.

[10] C. A. Bouman. (1997, April) Cluster: An unsupervised algorithm for modeling Gaussian mixtures. [Online]. Available: http://www.ece.purdue.edu/∼bouman

Metal artifact reduction based on the combined prior image

Yanbo Zhang, Xuanqin Mou

***Abstract*—Metallic implants introduce severe artifacts in CT images, which degrades the image quality. It is an effective method to reduce metal artifacts by replacing the metal affected projection with the forward projection of a prior image. How to find a good prior image is the key of this class methods, and numerous algorithms have been proposed to address this issue recently. In this work, by using image mutual correlation, pixels in the original reconstructed image or linear interpolation corrected image, which are less affected by artifacts, are selected to build a combined image. Thereafter, a better prior image is generated from the combined image by using tissue classification. The results of three patients' CT images show that the proposed method can reduce metal artifacts remarkably.**

***Index Terms*—Computed tomography, metal artifact reduction, prior image, mutual correlation.**

I. Introduction

METAL artifact reduction (MAR) is a major problem in x-ray computed tomography. Metallic implants can introduce bright and dark streaks and shadows in CT images, which degrades the image quality severely and become a major limiting factor in clinical diagnosis. During past three decades, various metal artifact reduction approaches have been proposed. However, there is still no robust solution to this issue and it remains a challenging problem.

The projections passing through metals are distorted by various errors such as severe beam hardening and noise [1]. As a result, many MAR methods treat metal affected projections as to be missing, and they are replaced by surrogate projections. Some methods complete the projection by using an interpolation scheme, e.g., linear interpolation [2] (denoted as LI-MAR), which is simple and with low computation cost. However, interpolation based MAR methods may introduce secondary artifacts.

Recently, a class of MAR methods, which complete the missing projection dataset by using forward projection of a prior image, are widely investigated. The information of the prior image is exploited to complete the projection, as a result, the method can get an excellent result if the prior image contains few artifacts and is close to the ground truth image. Therefore, how to find a good prior image is crucial in these forward projection based methods. Generally, a prior image is generated from the original reconstructed image or pre-corrected image. Seemeen Karimi et al. [3] obtained the prior image by segmenting regions of the original image. Metal artifacts regions were identified and then replaced with a constant soft tissue value. Bal and Spies [4] employed the k-means cluster technique to segment the adaptively filtered image into five classes. Prell et al. [5] segmented three dimensional interpolation corrected image into air, soft tissue and bone. Philips Healthcare recently developed a commercial orthopedic metal artifact reduction function (O-MAR) which produced the prior image from the original image [1]. Meyer et al. [6] produced prior image from different images depending on the strength of existing artifacts. The original image was chosen to generate prior image in the case of existing minor artifacts, otherwise, LI-MAR corrected image was selected instead. There are two main drawbacks for these methods. In some cases, there exist wrong tissue classification due to severe artifacts. Besides, pixel values in bone remain unchanged because they vary over a large range; as a result, the artifacts in bone remain. These factors result in generating poor prior image and finally lead to dissatisfactory correction performance.

All the above mentioned methods employ the information in only one image, while we try to make the best of the information in both the uncorrected original image and LI-MAR corrected image. The distributions and intensities of artifacts are different in the original image and LI-MAR corrected image, so the pixels containing fewer artifacts in the two images are selected to build a combined image, which is used to generate a good prior image. Thereafter, the forward projection of the prior image is used to complete the projection dataset and the corrected image is reconstructed using FBP.

II. Method

The main idea of our approach is to generate the prior image that is obtained from the combination of original image and LI-MAR corrected image. The proposed method is composed of three steps: Metal traces segmentation, prior image generation and projection completion followed by image reconstruction.

A. Metal traces segmentation

In the original reconstructed image, metals are segmented out based on thresholding [7]. Then the forward projection of the obtained metal only image is performed to get the metal traces, which specifies the projections affected by metals. These affected projections are replaced in the final step.

B. Prior image generation

1) Linear interpolation: For a given original sinogram $\mathbf{p}$, whose the i^{th} view j^{th} bin pixel is denoted as $p_{i,j}$. In each

This work was partly supported by the NSFC through Grant No. 61172163 and the Research Fund for the Doctoral Program of Higher Education of China through Grant No. 20110201110011.

Yanbo Zhang and Xuanqin Mou are with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China. (Email: yanbozhang007@163.com, xqmou@mail.xjtu.edu.cn).

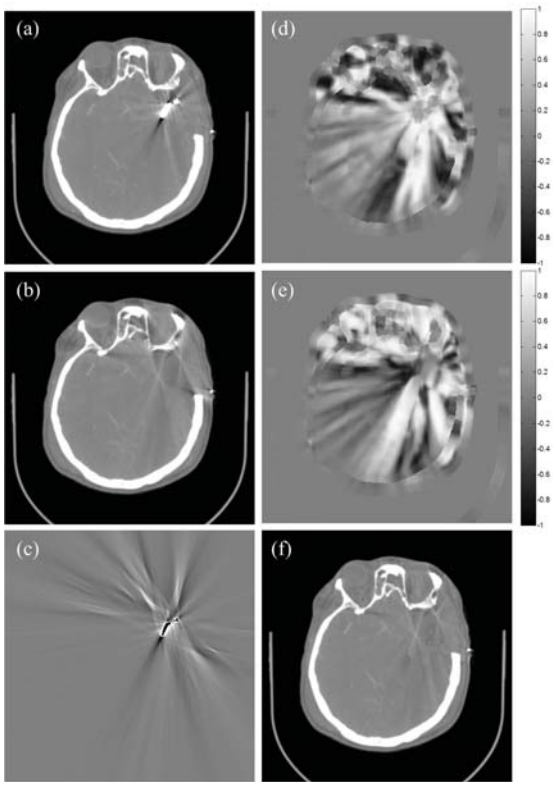


Fig. 1. Generation of the combined images of patient 1. (a) is the original image, (b) is LI-MAR corrected image, (c) is the difference between (a) and (b), where the metal pixels are excluded. (d) and (e) are the correlation maps of (a) and (b), respectively. (f) is the combined image. The display windows are (WL=0HU, WW=1500HU) for (a–c) and (f), and [-1, 1] for (d) and (e).

view of the sinogram, e.g. in the i^{th} projection view, if the projections $p_{i,k}$ and $p_{i,k+\Delta+1}$ are unaffected by metal, and the Δ projection pixels between them, $\{p_{i,j}|j \in [k+1, k+\Delta]\}$, are in metal trace, then the projections $\{p_{i,j}|j \in [k+1, k+\Delta]\}$ are replaced by the linear interpolation of $p_{i,k}$ and $p_{k+\Delta+1}$, which is stated as follows.

$$p_{i,j}^{LI} = p_{i,k} + \frac{p_{i,k+\Delta+1} - p_{i,k}}{\Delta + 1}(j - k). \tag{1}$$

Sometimes the unaffected projections are smoothed before interpolation in order to decrease the impact of noise. The unaffected projections remain unchanged in $\mathbf{p}^{LI}$. Then the LI-MAR corrected image is reconstructed using FBP.

2) Artifacts splitting: The original image $\mathbf{f}^O$ can be regarded as the ground truth image plus a metal artifacts image, and the LI-MAR corrected image $\mathbf{f}^{LI}$ can be treated as the ground truth image plus a LI-MAR artifacts image consisting of uncorrected metal artifacts and secondary artifacts. Therefore, the difference of these two reconstructed images, $\mathbf{f}^O - \mathbf{f}^{LI}$, represents the superposition of metal artifacts and the negative LI-MAR artifacts, called as artifacts superposition image and denoted as $\mathbf{f}^A$. Figure 1(c) gives the artifacts superposition image of a patient (patient 1) containing a metallic clip. It can be seen that all metal artifacts and LI-MAR artifacts are presented and there is no information of patient' tissues.

3) Correlation maps: Image mutual correlation [8] has been widely used to assess the degree of similarity of two images. The mutual correlation of two vectors $\mathbf{x}$ and $\mathbf{y}$ is calculated as follows:

$$C(\mathbf{x}, \mathbf{y}) = \frac{2\langle \mathbf{x}, \mathbf{y} \rangle + \epsilon}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \epsilon}. \tag{2}$$

where symbol $\langle \cdot, \cdot \rangle$ represents the inner product, and ϵ is a small positive constant to make sure that the denominator is not zero. The value range of $C(\cdot, \cdot)$ is $(-1, 1]$.

Generally, the distributions and intensities of metal artifacts and LI-MAR artifacts are different (e.g., Figs. 1(a) and 1(b)), so it is possible to build a combined image with fewer artifacts from the two images $\mathbf{f}^O$ and $\mathbf{f}^{LI}$. Specifically, for an arbitrary pixel position (i, j), the pixel value of the combined image is selected as $f_{i,j}^O$ or $f_{i,j}^{LI}$, depending on that whose CT number is less affected by artifacts. For this purposes, two reconstructed images $\mathbf{f}^O$ and $\mathbf{f}^{LI}$ as well as the artifacts superposition image $\mathbf{f}^A$ are divided into blocks (called image patches), and denoted as $\mathbf{b}_{i,j}^O$, $\mathbf{b}_{i,j}^{LI}$ and $\mathbf{b}_{i,j}^A$, respectively, where the subscript (i, j) is position index of the central pixel of the image patch. Each patch has 9×9 pixels. As known to us, CT number of soft tissue is around 0 HU; thus, in cases of heavy artifacts, the CT number is dominated by artifacts in soft tissue region in the reconstructed images. Therefore, in soft tissue region, correlation value between two corresponding patches in the reconstructed image and in artifacts superposition image describes the correlation degree of contained artifacts in this patch in the reconstructed image, which is calculated by Eq. 3.[1]

$$\begin{cases} C_{i,j}^O = C(\mathbf{b}_{i,j}^O, \mathbf{b}_{i,j}^A), \\ C_{i,j}^{LI} = -C(\mathbf{b}_{i,j}^{LI}, \mathbf{b}_{i,j}^A). \end{cases} \tag{3}$$

These correlation values obtained form Eq. 3 compose correlation maps $\mathbf{C}^O$ and $\mathbf{C}^{LI}$. Since the artifacts is continuous along the directions of streaks, correlation values in bone can be estimated by using neighbourhood interpolation. Then two correlation maps $\mathbf{C}^O$ and $\mathbf{C}^{LI}$ are obtained. Figures 1(d) and 1(e) are the two correlation maps of patient 1.

4) Generation of combined prior image: If the corresponding correlation value of $f_{i,j}^O$ is smaller than that of $f_{i,j}^{LI}$, then $f_{i,j}^O$ is likely to contain fewer artifacts and is selected to build the combined image; otherwise, $f_{i,j}^{LI}$ is chosen. Thus the combined image $\mathbf{f}^C$ is obtained according to Eq. 4. Figure 1(f) shows the combined image of patient 1, which contains lighter artifacts than Figs. 1(a) and 1(b).

$$f_{i,j}^C = \begin{cases} f_{i,j}^O, & \text{if } C_{i,j}^O < C_{i,j}^{LI} \\ f_{i,j}^{LI}, & others. \end{cases} \tag{4}$$

The prior image is obtained via tissue classification of the combined image [5]. The pixels with CT numbers larger than 200 HU are regarded as bone, which are unchanged; the pixels whose CT numbers are smaller than -600 HU, are assumed as air and set to -1000 HU; while the pixels with CT numbers between -600 HU and 200 HU are treated as soft tissue, which are uniformly set to 0 HU. Then the combined prior image is obtained.

[1]The artifacts superposition image is defined by $\mathbf{f}^A = \mathbf{f}^O - \mathbf{f}^{LI}$, so the negative sign is used in Eq. 3 to guarantee the positive correlation of LI-MAR induced artifacts between two corresponding patches in $\mathbf{f}^{LI}$ and $\mathbf{f}^O - \mathbf{f}^{LI}$.

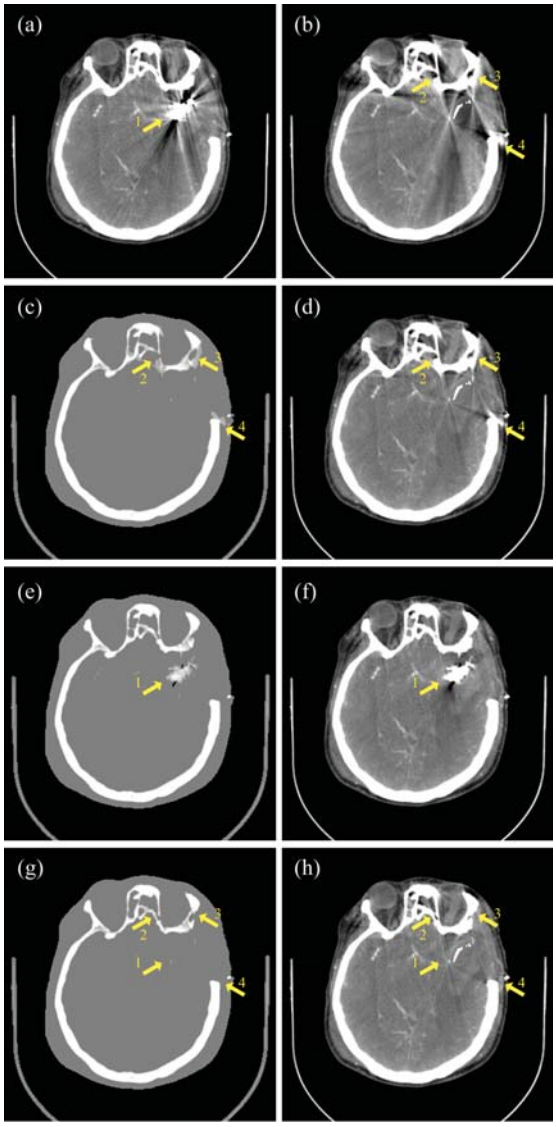


Fig. 2. Reconstructed images and prior images of patient 1 with a surgical clip. (a) is the original image, the corrected images are obtained by using (b) LI-MAR, (d) FP-MAR1, (f) FP-MAR2 and (h) the proposed method, respectively. (c), (e) and (g) are the prior images of (d), (f) and (h), respectively. The display windows are (WL=50HU,WW=400HU) for reconstructed images and (WL=0HU,WW=1500HU) for prior images.

C. Projection completion and image reconstruction

If the metal trace pixels are directly replaced with the corresponding projections obtained by the forward projecting of the combined prior image, it may lead to discontinuity at the boundary of the metal traces, which introduces new streak artifacts. So it is necessary to prevent generating the discontinuity in replacement. Similar to our previous work [9], we apply linear interpolation again to generate a continuous transition $\mathbf{p}^T$ between the forward projection $\mathbf{p}^C$ and the original sinogram $\mathbf{p}$. The metal affected projections are replaced with the sum of $\mathbf{p}^C$ and $\mathbf{p}^T$. In this way, the affected projections can be completed seamlessly. Thereafter, the corrected image is reconstructed using FBP, and the metal obtained in the first step is inserted into the corrected image.

III. Experimental Results

A. Experimental setup

In this study, we compare the performance of the proposed method with competing methods on the scanned datasets of three patients. A patient with a surgical clip (patient 1) was scanned on a Siemens SOMATOM Sensation 16 scanner CT using helical scanning geometry. The measurement of patient 1 was acquired with 1160 projection views over a rotation and 672 detector bins in a row. A patient with a dental filling (patient 2) and a patient with a hip prosthesis (patient 3) were scanned on a kV on-board imaging (OBI) system integrated in a $\mathrm{TrueBeam}^{\mathrm{TM}}$ medical linear accelerator (Varian Medical System, Palo Alto, CA). The projection datasets were acquired with 364 projection views over 200° in full-fan mode for patient 2, and 656 projection views over 360° in half-fan mode for patient 3, respectively, and their effective detector bins were 512. The matrix of reconstructed image is 512×512, corresponding pixel sizes are $0.776mm \times 0.776mm$ for patient 1 and patient 2, and $1mm \times 1mm$ for patient 3.

B. Results

In this paper, the forward projection based MAR methods which use the prior images generated from the original image and LI-MAR corrected image are denoted as FP-MAR1 and FP-MAR2, respectively. FP-MAR1 is the same to Prell et al. proposed method [5] except that the conventional LI-MAR is used instead of 3D LI-MAR. LI-MAR, FP-MAR1 and FP-MAR2 are implemented to compare with the proposed method.

Figure 2 shows the reconstructed images and the prior images of patient 1. The original image contains heavy metal artifacts in the vicinity of metallic clip as indicated by arrow 1 in Fig. 2(a). Thus the prior image obtained from the original image is not good enough due to wrong tissue classification as shown in Fig. 2(e), which results in severe artifacts in the corresponding reconstructed image (Fig. 2(f)). Figure 2(b) is the LI-MAR corrected image, which contains remarkable new artifacts as indicated by arrows 3 and 4, and the structure of bone is distorted highlighted by arrow 2. As a result, these artifacts and wrong structures are remained in the prior image (Fig. 2(c)) obtained from LI-MAR corrected image, leading to the similar artifacts and wrong structures as indicated by arrows in Fig. 2(d). By contrast, the combined prior image is obtained from Fig. 1(f). Therefore, there is no previously mentioned wrong tissue classification as indicated by arrows in Fig. 2(g), and the corrected image has no obvious artifacts (see Fig. 2(h)).

Figure 3 shows the reconstructed images of patient 2. There are obvious streak artifacts in both the original image and LI-MAR corrected image. FP-MAR1 can remove most of streaks except the one pointed out by the arrow in Fig. 3(c). In comparison, FP-MAR2 can suppress the streaks indicated by arrow better. Nevertheless, artifacts around the metal is remarkable in FP-MAR2 corrected image. The proposed method can reduce both streaks and artifacts around the metal greatly.

Figure 4 shows the reconstructed images of patient 3. There are bright and dark shadows in the original image, which can be reduced remarkably by LI-MAR. With the prior

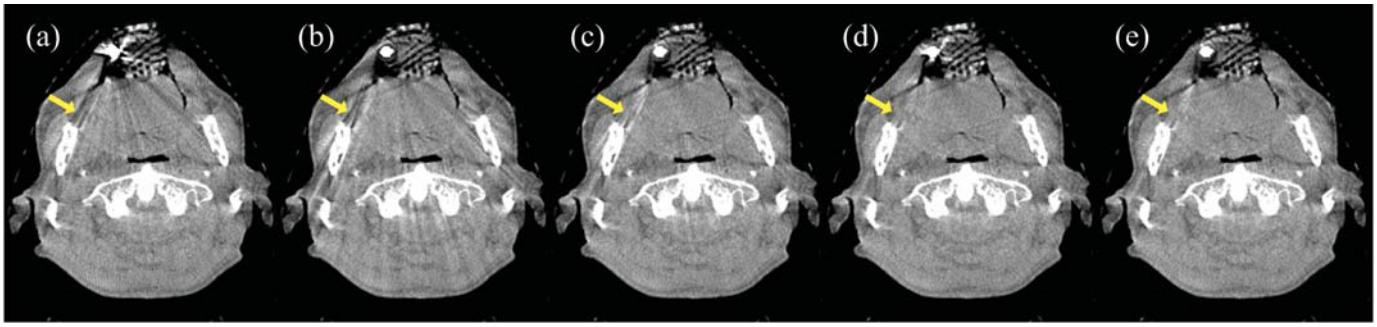


Fig. 3. Reconstructed images of patient 2 with a dental filling. (a) is the original image, the corrected images are obtained by using (b)LI-MAR, (c) FP-MAR1, (d) FP-MAR2 and (e) the proposed method, respectively. The display window is (WL=0HU,WW=750HU).

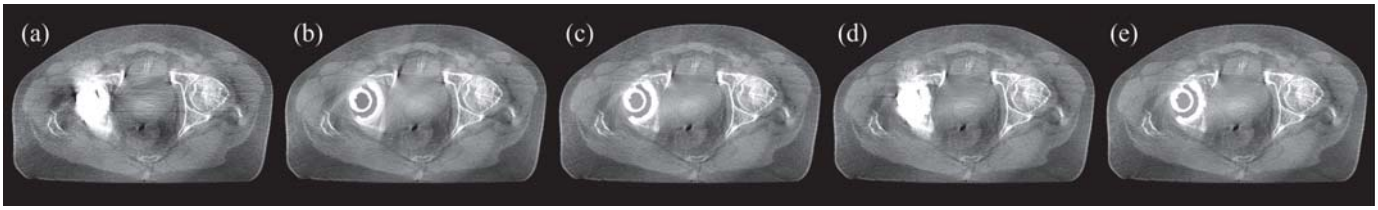


Fig. 4. Reconstructed images of patient 3 with a hip prosthesis. (a) is the original image, the corrected images are obtained by using (b)LI-MAR, (c) FP-MAR1, (d) FP-MAR2 and (e) the proposed method, respectively. The display window is (WL=0HU,WW=750HU).

image obtained from LI-MAR, FP-MAR1 can further alleviate artifacts. On the contrary, FP-MAR2 can hardly reduce bright artifacts around the hip prosthesis. The image corrected by the proposed method is similar to that by FP-MAR1, because the two prior images obtained from LI-MAR image and combined image are both good enough.

IV. Discussions and Conclusions

The distributions of artifacts in the original image and in LI-MAR corrected image are different, so the proposed method sufficiently exploits information of the pixels with fewer artifacts to compose a new image, which is used to generate a good prior image. As illustrated in the results of patient 1 and patient 2, images corrected by the proposed method are superior to that corrected by FP-MAR1 and FP-MAR2, because the combined prior image can avoid wrong tissue classification that appeared in the prior images of FP-MAR1 and FP-MAR2. For patient 3, since the prior image of FP-MAR1 has no wrong tissue classification and is superior to that of FP-MAR2, the prior image and corrected image of the proposed method are almost the same to that of FP-MAR1. Besides, for simplicity, LI-MAR method is adopted to generate the combined image in this work, while other MAR methods can also be used instead which will be our future work.

In conclusion, we introduce a new method to generate better prior image for the forward projection based metal artifact reduction method. By using image mutual correlation, pixels in the original image or linear interpolation corrected image, which are less affected by artifacts, are selected to build the combined image. Based on this image, a more accurate prior image can be obtained. The results demonstrate that the proposed method can achieve better artifacts removal performance than the competing methods. In the future, the developed method will be evaluated by clinicians to validate the clinical usefulness.

Acknowledgment

The authors would like to thank Dr. Hao Yan, Dr. Xun Jia and Dr. Steve B. Jiang for providing the datasets with the dental filling and the hip prosthesis, and thank Dr. Hengyong Yu for providing the dataset with the clip.

References

[1] Philips Healthcare, "Metal Artifact Reduction for Orthopedic Implants (O-MAR)," 2012. [Online]. Available: http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/CT/Metal-Artifact-Reduction-for-Orthopedic-Implants-(O-MAR)

[2] W. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," *Radiology*, vol. 164, no. 2, p. 576, 1987.

[3] S. Karimi, P. Cosman, C. Wald, and H. Martz, "Segmentation of artifacts and anatomy in CT metal artifact reduction," *Medical Physics*, vol. 39, no. 10, pp. 5857–5868, 2012.

[4] M. Bal and L. Spies, "Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering," *Medical Physics*, vol. 33, no. 8, pp. 2852–2859, 2006.

[5] D. Prell, Y. Kyriakou, M. Beister, and W. A. Kalender, "A novel forward projection-based metal artifact reduction method for flat-detector computed tomography," *Physics in Medicine and Biology*, vol. 54, no. 21, pp. 6575–6591, 2009.

[6] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, "Normalized metal artifact reduction (NMAR) in computed tomography," *Medical Physics*, vol. 37, no. 10, pp. 5482–5493, 2010.

[7] D. Prell, W. A. Kalender, and Y. Kyriakou, "Development, implementation and evaluation of a dedicated metal artefact reduction method for interventional flat-detector CT," *British Journal of Radiology*, vol. 83, no. 996, pp. 1052–62, 2010.

[8] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[9] Y. Zhang, H. Yan, X. Jia, J. Yang, S. B. Jiang, and X. Mou, "A hybrid metal artifact reduction algorithm for x-ray CT," *Medical Physics*, vol. 40, no. 4, p. 041910, 2013.

Bone artifact reduction in differential phase-contrast CT

Dieter Hahn, Pierre Thibault, Andreas Fehringer, Martin Bech, Peter B. Noël and Franz Pfeiffer

***Abstract*—In this work we present an image reconstruction technique for grating-based phase-contrast computed tomography. The main purpose is the reduction of artifacts, similar to metal artifacts known from conventional absorption imaging, caused by the presence of highly absorbing and scattering objects, e.g. bones. The method is based on a statistical iterative reconstruction algorithm utilizing maximum-a-posteriori (MAP) principles and integrating the statistical properties of the raw data as measured with a grating interferometer. As data processing is needed to calculate the final absorption, darkfield and differential phase-contrast signals from the raw projections, the measurement errors are propagated through the processing steps to obtain estimates of the statistical uncertainties to be used in the reconstruction procedure. To a good approximation these uncertainties follow a Gaussian probability distribution, leading to the formulation of a penalized log-likelihood cost function with a weighted least-squares data-fidelity term, complemented with one or more regularization terms incorporating prior knowledge of the reconstructed image. Information about the position of dense materials is gained from the absorption signal and utilized in the phase reconstruction. The technique is demonstrated on experimental data from a synchrotron X-ray source as well as from a laboratory setup. The results clearly show an improvement in image quality and are another step in proving the feasibility of phase-contrast techniques for clinical applications.**

I. Introduction

In the past few years grating-based X-ray phase-contrast imaging has increasingly attracted attention [1]–[3] because of its high potential to be introduced into clinical settings. In addition to the conventional attenuation information it provides a differential phase-contrast signal with a high sensitivity to small electron density variations, leading to an increased soft tissue contrast. This makes it a promising tool, for example, in early tumor detection. This high sensitivity, however, also causes one of its main challenges. Performing phase-contrast tomography of soft tissue regions in the presence of highly absorbing or scattering objects, e.g. bones, these structures will cause streaking artifacts in a conventional filtered backprojection reconstruction, the current standard reconstruction method [4]–[7]. These artifacts mainly arise due to high attenuation, phase wrapping and scattering of the bone material. Because in medical phase-contrast imaging these artifacts mostly appear near bones, we term them 'bone artifacts'. These bone artifacts drastically impair the diagnostic quality of the reconstructed images and could prevent the introduction of the technique into clinical routine. Luckily, with the increase in computational power available today, iterative reconstruction techniques have become a feasible alternative to the filtered backprojection algorithm. Their main advantage is the ability to use prior knowledge of the reconstructed image [8], i.e. fully modelling the physical and statistical processes involved in the imaging procedure.

D. Hahn, P. Thibault, A. Fehringer, M. Bech and F. Pfeiffer are with the Chair for Biomedical Physics and the Institute for Medical Engineering, Technische Universität München, Munich, Germany

P.B. Noël is with the Department of Radiology, Technische Universität München, Munich, Germany

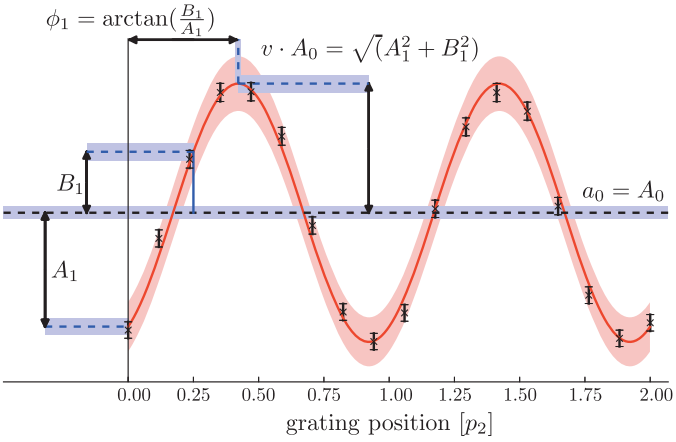


Fig. 1. X-ray intensity curve recorded during phase-stepping in a single pixel as a function of the position of the analyzer grating. Indicated are the Fourier coefficients (a_0, v and ϕ_1) and the new fit parameters (A_0, A_1 and B_1), as well as their statistical uncertainties.

In this work we propose a bone artifact reduction algorithm based on statistical iterative reconstruction (SIR) using maximum-a-posteriori principles. Further a simple alternative to the conventional Fourier data processing used to calculate the absorption and differential phase-contrast signals from the raw data is presented.

A grating interferometer basically consists of two gratings. The first so-called phase grating creates intensity modulations on the length scale of the grating period downstream of the beam. A phase-shifting object in the beam causes slight deviations of this pattern proportional to the angles of refraction caused by the object. A second absorbing grating with a period matched to that of the intensity pattern acts as an analyzer to resolve the fringes with a conventional X-ray detector. The pattern is scanned by moving one of the gratings perpendicular to the beam, thus recording an intensity curve for each individual pixel. Fourier analysis is typically used to retrieve attenuation, phase-contrast and darkfield signals from these curves. As we want to use the statistical uncertainties of the three signals to aid in the reconstruction, we use a weighted least-squares approach to directly fit a periodic curve to the

measurements.

$$S = \sum_j \frac{1}{\sigma_j^2} \left(I_j - A_0 - A_1 \cos(x_j) - B_1 \sin(x_j)\right)^2 . \quad (1)$$

A typical intensity curve and the relation between the conventional Fourier coefficients and the fitting parameters are shown in figure 1.

II. The Problem

With the statistical uncertainties of our projection data at hand, we have the first ingredient for our bone artifact reduction method. But before we go into the details of the reconstruction procedure we want to briefly discuss the causes that lead to the appearance of the bone artifacts.

1) The strong absorption of dense materials leads to photon starvation and loss of information.
2) Small-angle scattering inside these porous materials causes a loss of coherence and thus limits the ability to reliably determine the phase gradient.
3) Measurement of the phase gradient is intrinsically restricted to the interval $[-\pi, \pi]$. Gradients exceeding this range are wrapped back into this interval, causing an undefined value at this position, thus lowering its statistical significance. This phase-wrapping usually occurs at strong edges, where the differential phase shift is especially high, e.g. at the boundary between soft tissue and bones.

All of the above effects lead to a differential phase shift that is no longer uniquely defined at certain positions and thus does not represent reliable information for the tomographic reconstruction.

III. Methods and materials

To reduce the influence of these unreliable measurements we propose an algorithm that is based on statistical modeling of the processes involved in differential phase-contrast imaging. The probability distributions of the signals and their statistical uncertainties obtained after the processing procedure follow to a good approximation a Gaussian distribution. Gaussian behavior of a statistical model means that the log-likelihood in a Bayesian sense is written as a weighted sum of squared differences plus regularization terms embedding prior knowledge of the reconstructed image. The resulting cost function, called the penalized log-likelihood is to be minimized with respect to the reconstructed image ρ given the corresponding measured sinogram s and uncertainties σ:

$$\mathcal{L}(\rho) = \sum_{x,\theta} \frac{1}{(\sigma_x^\theta)^2} \left(s_x^\theta - \sum_j \partial_x A_{x,j}^\theta \rho_j \right)^2 + \sum_k \lambda_k R_k(\rho). \quad (2)$$

Here x is the coordinate of the pixels along a detector line, θ the projection angle and $\partial_x A_x^\theta$ the differential forward projection operator, projecting a tomogram into a differential sinogram, to take into account the differential nature of the phase-contrast projections. The R_k are the regularization terms with each having a strength factor λ_k. We minimize this cost function using a standard non-linear conjugate gradient algorithm.

But how can we use this algorithm to solve the problems stated above? Looking at the equation of the differential phase uncertainty unveils that it is inversely proportional to the darkfield signal (also shown theoretically in [9] and [10]), which is a direct measure for the small angle scattering in a sample. Another look at the phase uncertainty shows that it is also directly proportional to the absorption. Thus, using the phase uncertainty directly in the SIR should take care of problems 1. and 2. Problem 3 on the other hand requires more preparation. We know that phase-wrapping is most likely to occur at the boundaries between low and high absorbing regions and from an absorption reconstruction we can get information on the exact location of strongly absorbing objects. The absorption tomogram is thresholded and forwardprojected into a new sinogram. This sinogram is then differentiated along the detector coordinate direction to get an estimate of where strong edges occur. The weighting term $1/\sigma^2$ in the cost function is then modified using this 'mask' to decrease the weighting of pixels, which have a high probability of containing phase-wrapping.

In addition to the weighting, there is another part of the cost function that can be used to improve the reconstruction result, and this is the regularization.

In this work four regularization terms are utilized. Three of those are globally applied to the reconstructed image, the last one only works locally.

Quadratic regularization

The quadratic regularization term enforces a smooth reconstruction by quadratically penalizing the difference between a voxel and its surrounding neighbors. In three dimensions $\mathcal{N}_j$ is the neighborhood of voxel j and $w_{i,j}$ is a weighting factor taking into account the different distances of adjacent and diagonal neighbors.

$$R_Q(\rho) = \sum_j \sum_{i \in \mathcal{N}_j} w_{i,j} \left(\rho_j - \rho_i\right)^2 \quad (3)$$

Huber regularization

The Huber regularization term has an additional parameter that can be tuned. Differences larger than this threshold γ are penalized only linearly, while for differences smaller than γ the usual quadratic penalty is applied. This term enforces smoothness for already uniform regions, but preserves edges found in the image.

$$R_H = \sum_j \sum_{i \in \mathcal{N}_j} \begin{cases} \frac{(\rho_j - \rho_i)^2}{2\gamma^2} & \text{for } |\rho_j - \rho_i| < \gamma \\ \frac{|\rho_j - \rho_i| - \gamma/2}{\gamma} & \text{for } |\rho_j - \rho_i| > \gamma \end{cases} \quad (4)$$

Mean regularization

This term penalizes large differences of voxel values to the mean of their neighborhood. The size of the neighborhood can

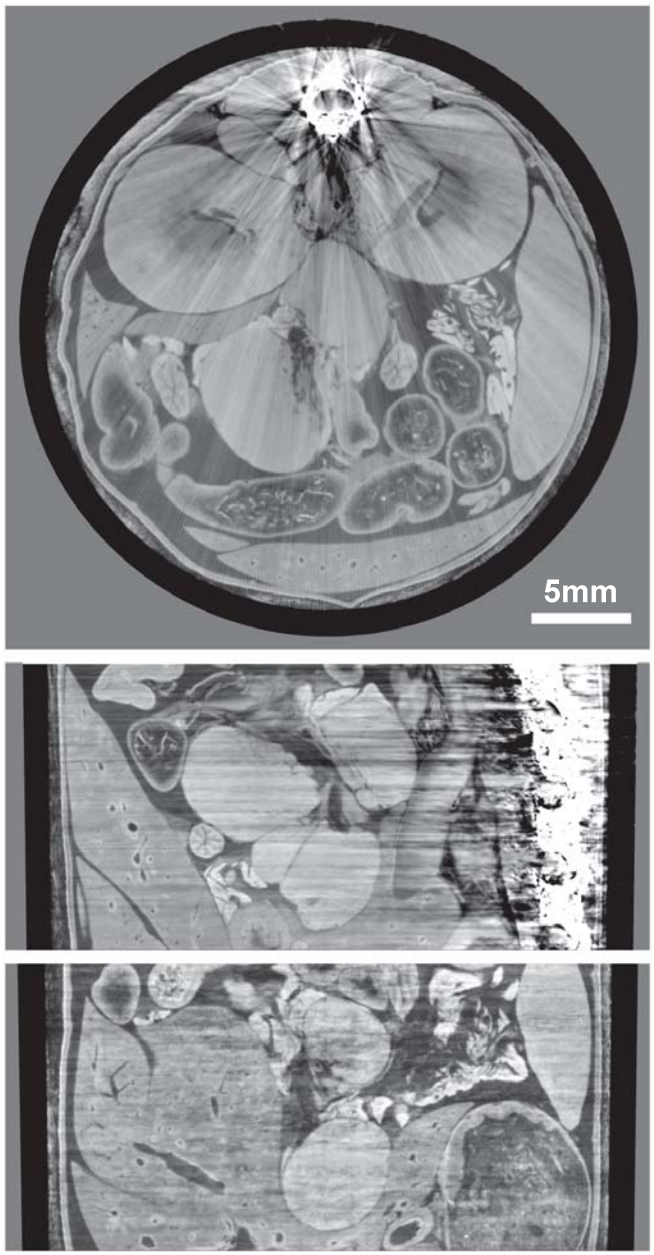


Fig. 2. Axial (top), sagittal (middle) and coronal (bottom) cuts through the X-ray phase-contrast CT of an ex-vivo mouse using filtered backprojection. The streaking artifacts caused by the spine are clearly visible in the axial view and also show an influence on the perpendicular views as strong pixel value variations. More details on the experiment can be found in [11]

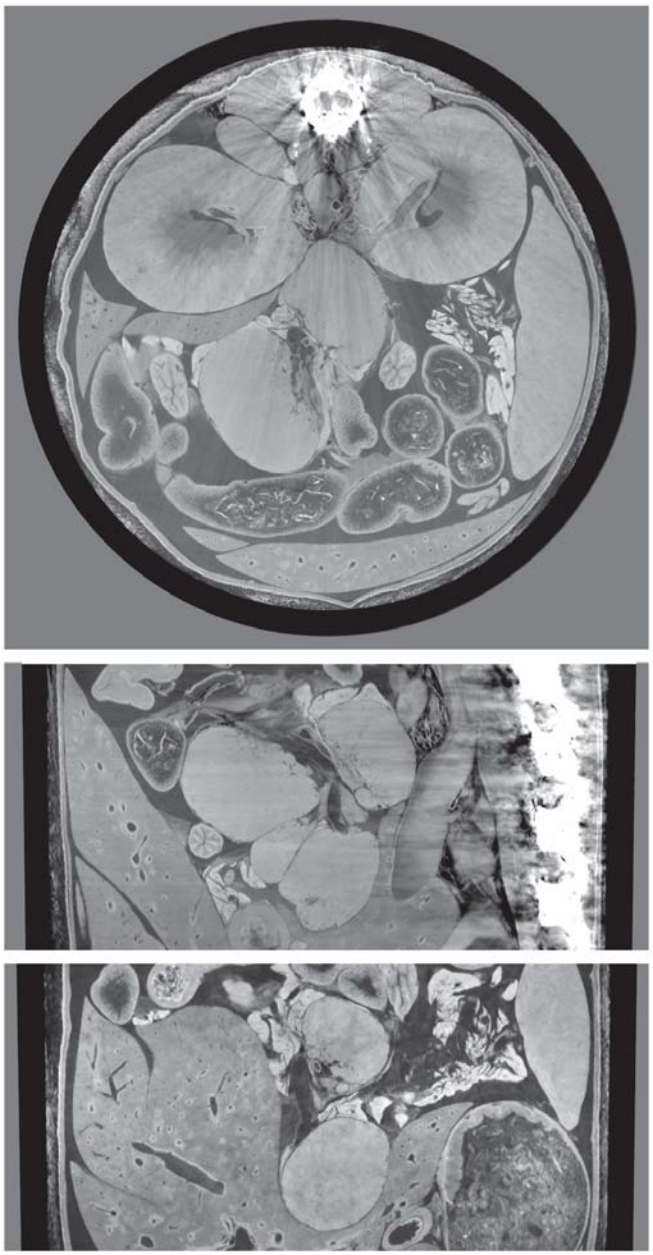


Fig. 3. Axial (top), sagittal (middle) and coronal (bottom) cuts through the X-ray phase-contrast CT of an ex-vivo mouse measurement using the proposed SIR algorithm. Compared to the filtered backprojection reconstruction in figure 2, the streaking artifacts are clearly reduced in the axial view and the stripes and structural artifacts in the perpendicular views are also suppressed.

be chosen freely. The purpose of this term is to enforce the reduction of long range deviations.

$$R_M = \sum_j \left(\rho_j - \langle\rho\rangle_{\mathcal{N}_j}\right)^2 \tag{5}$$

Bone regularization

This term is only applied to pixels that contain dense material using the mask b generated from an absorption reconstruction. The value of affected voxels is forced to be similar to absorption values scaled by a factor c. This factor is ideally chosen as δ/μ for the underlying material to enforce 'correct' phase values.

$$R_B(\rho) = \sum_j b_j \left(\rho_j - c \cdot a\right)^2 \tag{6}$$

IV. Results

The algorithm's ability to reduce bone artifacts is demonstrated on two experimental datasets. The first one is an ex-vivo phase-contrast CT scan of a mouse measured at the European Synchrotron Radiation Facility in Grenoble. Here

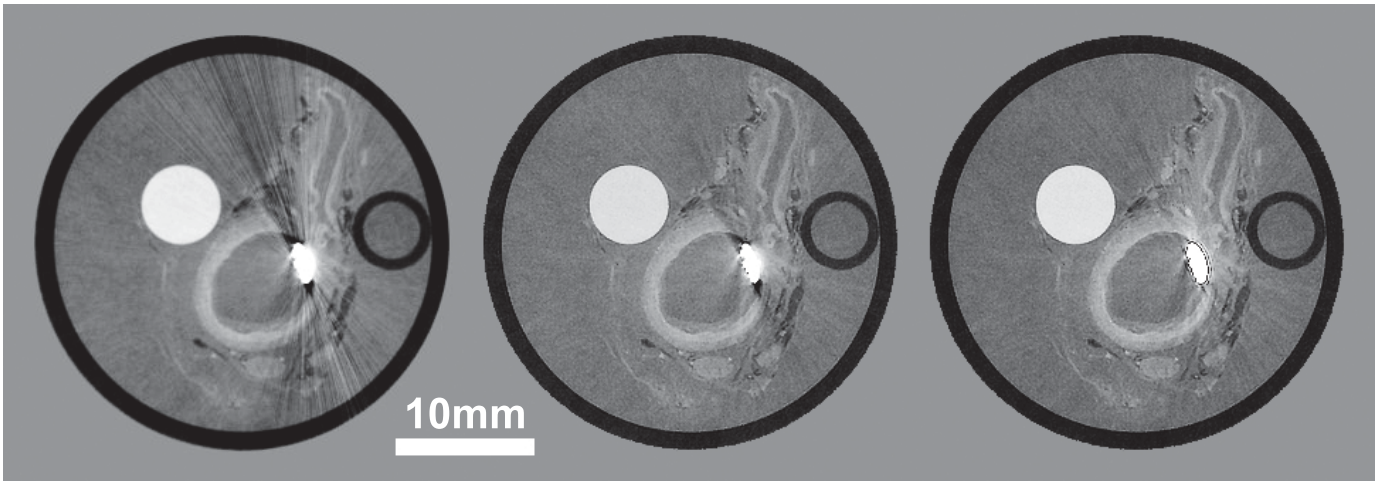


Fig. 4. Reconstructions of a carotid artery measured at a laboratory setup. The artery contains calcified plaque causing streaking artifacts as seen in the FBP reconstruction on the left. The middle plate shows a reconstruction using the proposed algorithm only utilizing statistical weighting and Huber regularization. The streaking artifacts are suppressed but there are dark shadows left around the calcifications. These can be removed by further constraining the reconstruction with the bone regularizer as seen in the image on the right.

we present the results of reconstructions performed using the conventional filtered backprojection (figure 2) and the proposed method (figure 3). Shown in the figures are an axial slice (top) and sagittal (middle) and coronal (bottom) cuts through the center of the volume. The FBP reconstruction exhibits strong streaking artifacts that manifest as vertical lines in the sagittal view and a noise-like structure in the coronal view obstructing most of the fine detail. While there are still artifacts left in the reconstruction with the presented method, their intensity and extension is drastically reduced, making the detection of small details a lot more feasible. This becomes most apparent in the coronal view, where the influence of the artifacts is no longer visible. In addition to the artifact reduction, the iterative reconstruction results in sharper images with more local contrast.

To demonstrate the general applicability of the method, it was also used on a second dataset measured at a laboratory setup with a conventional rotating anode X-ray source. The sample is an excised carotid artery with calcified plaque. The results are shown in figure 4 with the FBP reconstruction on the left, an SIR reconstruction without the bone regularizer in the middle and an SIR reconstruction with bone regularizer on the right. It is apparent from the left image, that these calcifications cause artifacts impairing the detectability of details in the underlying tissue. Using only the information on the statistical uncertainties in the iterative reconstruction, the streaking artifacts are suppressed, leaving only dark shadows around the calcification, as seen in the middle plate. These can also be removed by further constraining the reconstruction with the help of the bone regularizer, shown in the right image, further increasing detail visibility in the vicinity of the calcification.

V. Conclusions

We demonstrated a new method for reducing the influence of strongly absorbing and scattering objects on the reconstruction of differential phase-contrast tomography data, decreasing the negative impact the resulting artifacts have on medical diagnosis. This work thus expands the potential use cases of grating-based differential phase contrast for biomedical imaging from currently mostly soft tissue applications to a much wider range. With our method's ability to decrease the influence of bone artifacts X-ray phase-contrast imaging becomes one step closer to being clinically feasible.

Acknowledgements

We acknowledge the help of Irene Zanette, Arne Tapfer, Timm Weitkamp during the synchrotron measurements, and Marian Willner, Alexander Hipp, Julia Herzen during the x-ray tube measurements.

References

[1] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, *Nature Physics*, vol. 2, no. 4, pp. 258–261, Mar. 2006.
[2] A. Momose, W. Yashiro, Y. Takeda, Y. Suzuki, and T. Hattori, *Japanese Journal of Applied Physics*, vol. 45, no. 6A, pp. 5254–5262, 2006.
[3] T. Weitkamp, A. Diaz, C. David, F. Pfeiffer, M. Stampanoni, P. Cloetens, and E. Ziegler, *Optics Express*, vol. 13, no. 16, pp. 6296–6304, 2005.
[4] F. Pfeiffer, O. Bunk, C. David, M. Bech, G. Le Duc, A. Bravin, and P. Cloetens, *Physics in Medicine and Biology*, vol. 52, no. 23, pp. 6923–30, Dec. 2007.
[5] F. Pfeiffer, C. Kottler, O. Bunk, and C. David, *Physical Review Letters*, vol. 98, no. 10, pp. 108 105–108 108, Mar. 2007.
[6] M. Bech, T. H. Jensen, R. Feidenhans'l, O. Bunk, C. David, and F. Pfeiffer, *Physics in medicine and biology*, vol. 54, no. 9, pp. 2747–53, May 2009.
[7] T. Donath, F. Pfeiffer, O. Bunk, C. Grünzweig, E. Hempel, S. Popescu, P. Vock, and C. David, *Investigative Radiology*, vol. 45, no. 7, pp. 445–452, 2010.
[8] J. A. Fessler, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, M. Sonka and J. M. Fitzpatrick, Eds. Bellingham: SPIE, 2000.
[9] K. J. Engel, D. Geller, T. Köhler, G. Martens, S. Schusser, G. Vogtmeier, and E. Rössl, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 648, pp. 202–207, 2010.
[10] G.-H. Chen, J. Zambelli, K. Li, N. Bevins, and Z. Qi, *Medical Physics*, vol. 38, no. 2, p. 584, 2011.
[11] A. Tapfer et al., *PLOSOne*, vol. 8, no. 3, p. e58439, 2013.

A New Method for Windmill Artifact Reduction

Johan Sunnegårdh, Karl Stierstorfer, Siemens Healthcare, Germany

***Abstract*—A common problem in helical cone-beam Computed Tomography (CT) is windmill artifacts. We propose a new method for reduction of such artifacts, based on estimation and subsequent subtraction of the artifacts. Experiments on different phantoms demonstrate the reduction of windmill artifacts, and the surprising side effect of enhancement of high contrast edges. In the final section, we discuss how certain low-pass filter parameters affect the result, as well as possible improvements of the method.**

I. Introduction

WINDMILL artifacts typically occur in helical cone-beam CT. In reconstructed axial images, the artifacts appear as alternating dark and brigth areas emanating from objects containing sharp edges in the z-direction. The cause of these artifacts is aliasing of high frequencies in the z-direction [1]. To see this, we assume a point focus, and a detector that has a sampling distance S projected to the iso-center in the z-direction. Typically, a CT detector has an aperture that as approximately as large as the samling distance. Thus, the Modulation Transfer Function (MTF) of the detector aperture is a sinc function with its first zero at $1/S$, *i.e.*, twice the Nyquist frequency of the detector in the z-direction. Since this function only falls to approximately 0.6 of its maximum at the Nyquist frequency, its anti-aliasing properites are insufficient.

A hardware solution to the windmill artifact problem is the z-Flying Focal Spot (z-FFS) [2]. By rapidly deflecting the electron beam so that the focal spot jumps between two different z-positions, an effective detector sampling distance of $S/2$ is obtained. This corresponds to the increased Nyquist freqency of $1/S$, coinciding with the first zero of the detector aperture MTF. Thus, the most severe aliasing distortion is avoided. While very effectively suppressing the windmill artifacts already at the sampling stage, z-FFS has the disadvantage that it increases the complexity and cost of the x-ray tube.

Since windmill artfacts are mainly caused by high frequencies in the z-direction, a straightforward way to reduce the artifacts is to apply a low-pass filter in the z-direction. The artifact reduction would then come to the cost of reduced cross-plane resolution.

Several software approaches for windmill artifact reduction have been proposed. One approach is to adaptively mix sharp and soft contributions, either in the projection domain [3], [4], or in the image domain [5]. Another approach is to upsample data using shifted linear interpolation prior to backprojection [6]. Recently, Brown and Zabic [7] proposed the Windmill Artifact Reduction Processing (WARP) algorithm, which employs Total Variation (TV) minimization in different directions, to extract and subtract windmill artifacts. Although all of these methods are better than simple low-pass filtering, they all seem to contain steps that potentially might compromise z-resolution.

In the following, we propose a three step algorithm for reduction of windmill artifacts. First, an edge recovery operator is applied to the image in order to recover the strucures that cause windmill artifacts, *i.e.*, high contrast edges orthogonal to the z-axis. Second, windmill artifacts are estimated by forward projecting and reconstructing the edge image. In the final step, the estimated windmill artifacts are subtracted from the original reconstruction.

II. Materials and methods

A. Windmill artifact reduction algorithm

We use a vector/matrix notation for representing images and data, and linear operations. The vector $\boldsymbol{p} \in \mathbb{R}^M$ denotes projection data and the vector $\boldsymbol{f} \in \mathbb{R}^N$ denotes images. The matrix $\boldsymbol{Q} \in \mathbb{R}^{N\times M}$ denotes a reconstruction operator, and the matrix $\boldsymbol{P} \in \mathbb{R}^{M\times N}$ denotes a forward projector. In our case, $\boldsymbol{Q}$ is the Weighted Filtered Backprojection (WFBP) [8], and $\boldsymbol{P}$ is the forward projection by Joseph [9]. However, most ideas presented here should be valid for other linear reconstruction methods and forward projectors.

The new method is based on the assumptions that windmill artifacts mainly are caused by high-contrast edges orthogonal to the z-axis, and that there is a nonlinear operator $F : \mathbb{R}^N \to \mathbb{R}^N$ that approximately can recover these edges from reconstructed images. Of course, these assumptions are not correct in all cases. There are high-frequency non-edge structures that may cause windmill artifacts, and there are nonregular edges that cannot accurately be approximated from reconstructed images. However, in many situations this assumption seems to be satisfied.

In the following, we consider forward projection followed by reconstruction to be the sum of a low-pass filter $\boldsymbol{L} \in \mathbb{R}^{N\times N}$ and a linear operator $\boldsymbol{A} \in \mathrm{R}^{N\times N}$ that causes artifacts (streaks, cone artifacts, and windmill artifacts):

$$\boldsymbol{QP} = \boldsymbol{L} + \boldsymbol{A}. \tag{1}$$

The operator $\boldsymbol{L}$ reflects low-pass filtering caused by interpolations in the forward projection and backprojection. Although these effects are normally spatially variant, we will approximate $\boldsymbol{L}$ with a spatially invariant filter, implying that any spatial deviations are considered to be artifacts, *i.e.* part of $\boldsymbol{A}$.

As a preparation, a reconstructed image $\tilde{\boldsymbol{f}} = \boldsymbol{Qp}$ containing windmill artifacts and other artifacts is calculated. The correc-

tion step is now given by

$$\boldsymbol{f} = \tilde{\boldsymbol{f}} - \underbrace{(\boldsymbol{QP} - \boldsymbol{L})}_{\text{artifact operator}} \underbrace{F(\tilde{\boldsymbol{f}})}_{\text{edges}}. \quad (2)$$

First, the edge recovery operator F is used for calculating an approximate edge image $F(\tilde{\boldsymbol{f}})$. Then, the artifact operator $\boldsymbol{A} = \boldsymbol{QP} - \boldsymbol{L}$ is applied in order to obtain estimates of the artifacts. In the final step, the estimated artifacts are subtracted from the original image $\tilde{\boldsymbol{f}}$.

B. The edge recovery operator $F : \mathbb{R}^N \to \mathbb{R}^N$

The main goal of F is to undo the smoothing that occurs during data acquisition and reconstruction, assuming that the edges originally were as sharp as step functions. Preferably, low contrast regions should become smoother, since this would reduce any impact the method might have on low contrast structures. There are probably many suitable approaches for solving this problem. In this paper we have employed a simple modified median filter approach.

For each output voxel f_i, a set Ω_i of neighborhood input voxels is considered. Algorithm 1 below contains a detailed pseudocode which is processed for each output pixel. The operator is a median filter in low-contrast regions. In high-contrast regions, *i.e.*, where the difference between maximum and minimum values in Ω_i is larger than a threshold T, the input voxel value is mapped to a value close to the minimum or close to the maximum depending on which is closer.

Algorithm 1 The inner loop of the edge recovery algorithm

parameters: $T \in (0, +\infty)$, $m \in [0, 1]$, and $\gamma \in (0, 1]$
$f_{i,\min} \leftarrow \min(\Omega_i)$
$f_{i,\max} \leftarrow \max(\Omega_i)$
$f_{i,\text{med}} \leftarrow \text{median}(\Omega_i)$
$f_{i,\text{contr}} \leftarrow f_{i,\max} - f_{i,\min}$
$f_{i,\text{marg}} \leftarrow m \cdot f_{i,\text{contr}}$
if $f_{i,\text{contr}} < T$ **then**
  return $f_{i,\text{med}}$
else
  $L \leftarrow f_{i,\min} + f_{i,\text{marg}}$
  $H \leftarrow f_{i,\max} - f_{i,\text{marg}}$
  if $f_i < L \vee f_i > H$ **then**
    return $f_{i,\text{med}}$
  else
    return $g_{L,H,\gamma}(f_i)$
  end if
end if

The function $g_{L,H,\gamma} : \mathbb{R} \to \mathbb{R}$ describes a transition between L and H, and is defined by

$$\begin{aligned} g_{L,H,\gamma}(x) &= \frac{\text{sign}(x')\,|x'|^{\gamma}\,(H-L)}{2} + \frac{L+H}{2} \\ x' &= 2\frac{(x - \frac{L+H}{2})}{H-L}. \end{aligned} \quad (3)$$

For $\gamma = 1$, the transition is linear, and for γ close to zero, the transition resembles a step function.

C. Relation to Regularized Iterative Weighted Filtered Backprojection (RIWFBP)

If F is the identity mapping, *i.e.* $F = I$, the correction equation (2) is identical to Regularized Iterative Weighted Filtered Backprojection (RIWFBP) [10] with a step length of $\alpha = 1$. To see this, let $F = I$ and formulate equation (2) as an iterative update step

$$\boldsymbol{f}_{k+1} = \tilde{\boldsymbol{f}} - (\boldsymbol{QP} - \boldsymbol{L})\boldsymbol{f}_k. \quad (4)$$

This update step can be stabilized by constructing each new iterate as a convex combination of the old iterate and the right hand side of the above equation, *i.e.*,

$$\boldsymbol{f}_{k+1} = \alpha\left(\tilde{\boldsymbol{f}} - (\boldsymbol{QP} - \boldsymbol{L})\boldsymbol{f}_k\right) + (1-\alpha)\boldsymbol{f_k}. \quad (5)$$

By substituting $\boldsymbol{L}$ with the linear regularization operator $\boldsymbol{R}_\beta \triangleq \boldsymbol{I} - \boldsymbol{L}$, we arrive at

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_k - \alpha(\boldsymbol{QP}\boldsymbol{f}_k - \tilde{\boldsymbol{f}} + \boldsymbol{R}_\beta \boldsymbol{f}_k), \quad (6)$$

which is the image based version of RIWFBP. By linearity, this update equation is equivalent to the projection domain formulation

$$\boldsymbol{f}_{k+1} = \boldsymbol{f}_k - \alpha(\boldsymbol{Q}(\boldsymbol{P}\boldsymbol{f}_k - \boldsymbol{p}) + \boldsymbol{R}_\beta \boldsymbol{f}_k). \quad (7)$$

The new formulation of the iteration shows that if the RIWFBP regularization operator $\boldsymbol{R}_\beta$ is chosen as $\boldsymbol{R}_\beta = \boldsymbol{I} - \boldsymbol{L}$, where $\boldsymbol{L}$ models the low-pass caused by the $\boldsymbol{QP}$ operator, undesired modification of the MTF caused by the RIWFBP is avoided. This is also valid for the converged result, since

$$\boldsymbol{f}_\infty = (\boldsymbol{QP} + \boldsymbol{R}_\beta)^{-1}\boldsymbol{Qp} = (\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{Qp}. \quad (8)$$

D. Experiments

To examine the effectiveness of the proposed method, two mathematical phantoms: the Turbell clock phantom [11] and the Forbild head phantom [12], and two physical phantoms: an anthropomorphic head phantom and the Catphan® high resolution module CTP 528 were used. The clock and head phantoms were used to study reduction of artifacts, and possible side-effects, and the high resolution module was used to study cross-plane resolution. Scanning and reconstruction parameters are listed in Table I. Pitch factors for the different data sets were 1.35 for the clock phantom, 0.36 for the Forbild head phantom, 0.3 for the anthropomorphic head phantom, and 0.55 for the CTP528 module. The CTP528 module was rotated so that the 7 line pair per centimeter pattern was oriented along the z-axis.

In all cases, the operator $\boldsymbol{L}$ was a Gaussian filter with standard deviation coefficients $(\sigma_x = 0.35, \sigma_y = 0.35, \sigma_z = 0.9)$ (voxel units). These coefficients were empirically determined by studying the difference $(\boldsymbol{QP} - \boldsymbol{L})F(\tilde{\boldsymbol{f}})$ for different choices. In order to avoid loss of resolution due to spatial variances in the MTF of $\boldsymbol{QP}$, the standard deviations were set to be slightly lower than necessary. This leads to a slight enhancement of high contrast edges, but should not noticably affect low contrast structures.

TABLE I. SCANNING AND RECONSTRUCTION PARAMETERS.

Number of channels	N_{ch}	736
Number of detector rows	N_{rows}	32 (Clock:64)
Number of projections/turn	$N_{\mathrm{proj},2\pi}$	1152
Scanning Field of View (SFOV)	R_{SFOV}	500mm
Slice width	S	1.2mm
Voxel volume in-plane sampling distance	Δ_{xy}	0.5mm
Voxel volume cross-plane sampling distance	Δ_z	0.3mm
Median filter dimensions	k_x, k_y, k_z	3,3,11
Contrast threshold	T	300HU
Relative contrast margin	m	0.05
Transition function shape factor	γ	0.25
WFBP redundancy weight	Q	0.8

To model the detector size, each detector element measurement was calculated as a mean of 2×2 line integrals. Although this does not reflect the actual physical data acquisition, it preserves the linearity of the forward projection model.

In order to simulate the data acquisition accurately, forward projection was done to the original (non-rebinned) cone-beam scanner geometry instead of the semi-parallel (rebinned) geometry suggested in [10].

For the initial reconstructions, *i.e.*, $\tilde{\boldsymbol{f}} = \boldsymbol{Qp}$, clinical reconstruction kernels (rampfilters) were used: "B31" for the clock and head phantoms and "B60" for the CTP528 module. In the correction step, the Shepp-Logan kernel [13] was used.

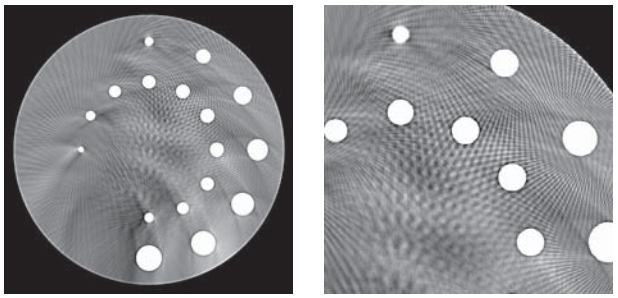
III. RESULTS

Fig. 1 a) and b) show a normal WFBP reconstruction of the clock phantom. Since the cone angle is relatively high ($\pm3.7°$), not only windmill artifacts, but also low-frequency cone-artifacts due to the non-exactness of the WFBP reconstruction appear.

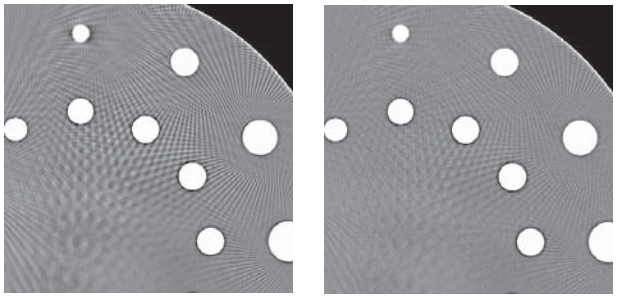
Fig. 1 c) shows the result after one RIWFBP iteration, or equivalently, after the correction step with $F = I$. In this result, most low-frequency cone-artifacts have been suppressed, but the windmill artifacts remain.

Fig. 1 d) shows the result of the proposed method. Here, also the windmills are less pronounced, although some higher frequencies still distort the image. The standard deviation in the center of the phantom is 8.7HU for the initial WFBP reconstruction, 5.0HU for the case where $F = I$, and 3.2HU for the new method. Apart from the reduced windmill artifacts, the new method also improves the reconstruction of the upper 12 o'clock sphere, which displays a cone-artifact in Fig. 1 b) that has not been perfectly suppressed in Fig. 1 c).

Fig. 2 a) - c) show the corresponding results for the Forbild head phantom. This phantom poses a greater challenge for the edge recovery operator, since it contains densely packed edges with high curvature in the inner ear region. As in the clock phantom case, the WFBP reconstruction again shows a mixture of cone artifacts and windmill artifacts. Correction without the edge recovery operator slightly improves both types of artifacts, but the improvement is more pronounced with the new method. The difference image in Fig. 2 d) verifies that while windmill artifacts are reduced, low-contrast structures remain unchanged. The obvious side effect is an enhancement of high-contrast edges.

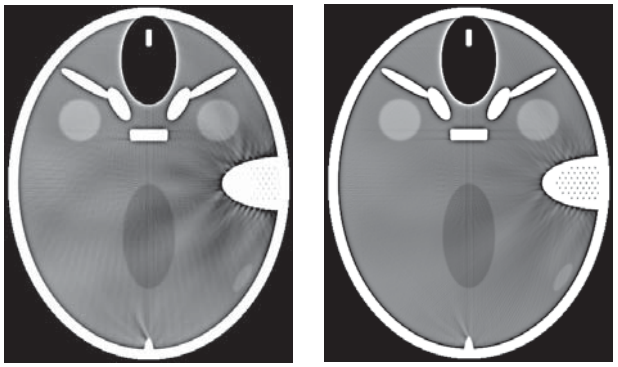


a) WFBP ($\tilde{\boldsymbol{f}}$) b) WFBP ($\tilde{\boldsymbol{f}}$), Zoomed

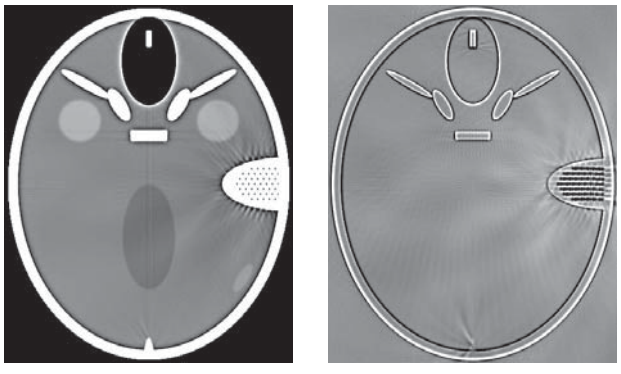


c) RIWFBP $\boldsymbol{f}_1$, ($F = I$) d) New method

Fig. 1. Turbell clock phantom reconstructions. Greyscale window (C:1000HU, W:80HU).

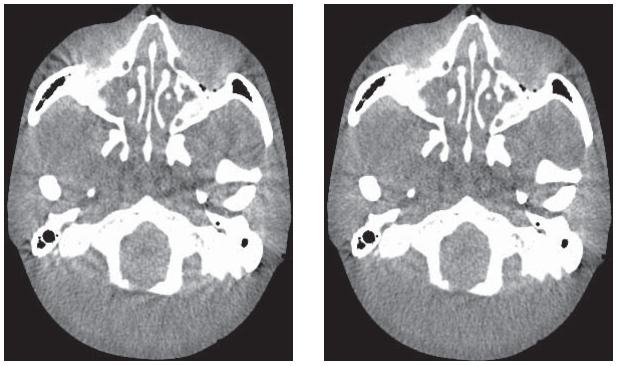


a) WFBP ($\tilde{\boldsymbol{f}}$) b) RIWFBP $\boldsymbol{f}_1$, ($F = I$)



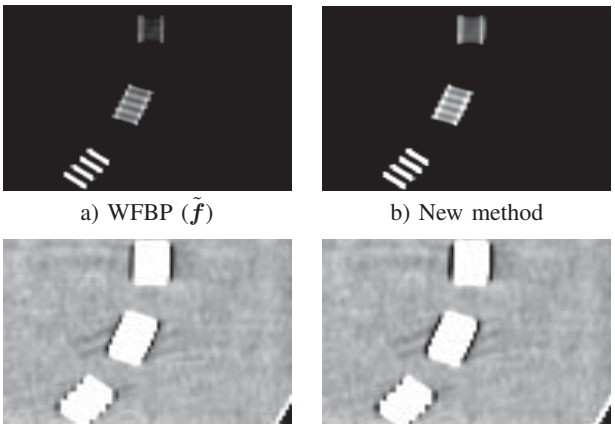
c) New method d) (c)-(a)

Fig. 2. Forbild head phantom reconstructions. Greyscale window (C:1050HU, W:50HU).



a) WFBP ($\tilde{\boldsymbol{f}}$) d) New method

Fig. 3. Anthropomorphic head phantom reconstructions. Greyscale window (C:1080HU, W:150HU).



a) WFBP ($\tilde{\boldsymbol{f}}$) b) New method

c) WFBP ($\tilde{\boldsymbol{f}}$) d) New method

Fig. 4. Coronal slices of reconstruction of the Catphan® CTP528 module. The z-axis points in the vertical direction, and the 7 line pairs per centemeter pattern is perfectly orthogonal to the z-axis. Greyscale window (C:2600HU, W:500HU) for a) and b), and (C:1100HU, W:300HU) for c) and d).

Reconstructions of the anthropomorphic head phantom are shown in Fig. 3. Here, the WFBP reconstruction show almost no cone-artifacts, but relatively strong windmill artifacts, which are much less pronunced in the corrected image. The difference image (not shown) shows windmill artifacts and high contrast edges similar to Fig. 2 d).

Fig. 4 shows coronal images of the CTP528 module. Here, enhanced edges can be observed, but there are no signs of reduced spatial resolution. Since the high frequency pattern causing windmill artifacts cannot be recovered by the edge recovery operator, no reduction of windmill artifacts can be seen in Fig. 4 c) and d).

IV. Discussion and Conclusion

As shown in the previous section, the proposed method reduces windmill artifacts without losing resolution or changing the noise texture in the images. Furthermore, the new method removes high frequency cone artifacts that are not properly suppressed by the RIWFBP.

One side effect is a change of edge responses: by constructing $\boldsymbol{L}$ to be slightly sharper than $\boldsymbol{QP}$, high contrast edges are slightly enhanced, and by constructing $\boldsymbol{L}$ to be softer than $\boldsymbol{QP}$, a blurring of high contrast edges will be seen. Another potential side effect introduction of new artifacts due to erroneous recovery of edges, or a bad system model $\boldsymbol{P}$. However, this effect was not observed in the experiments.

Clearly, further investigations are needed to understand why windmill artifacts remain in the corrected images. Many possible explanations exist, for instance, inaccurate recovery of edges, inaccurate system model (means calculated in attenuation domain instead of intensity domain, focal spot not modeled, reading integration not modeled), or mismatch between the reconstruction kernel used for initialization and the kernel used in the correction step.

References

[1] M. Silver, K. Taguchi, I. Hein, B. Chiang, M. Kazama, and I. Mori, "Windmill artifact in multi-slice helical CT," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5032 III, 2003, pp. 1918–1927.

[2] T. Flohr, K. Stierstorfer, S. Ulzheimer, H. Bruder, A. N. Primak, and C. H. McCollough, "Image reconstruction and image quality evaluation for a 64-slice CT scanner with z-flying focal spot," *Medical Physics*, vol. 32, no. 8, pp. 2536–2547, 2005.

[3] J. Hsieh, "Adaptive Interpolation Approach for Multi-slice Helical CT Reconstruction," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5032, 2003.

[4] G. Shechter, "Windmill Artifact Reduction in Multi-slice CT Reconstruction," Patent EP1 714 254B1, December, 2007.

[5] I. A. Hein, K. T. Taguchi, and S. M. D., "Method for helical windmill artifact reduction with noise restoration for helical multislice CT," U.S. Patent US7 623 691B2, November, 2009.

[6] A. A. Zamyatin, I. A. Hein, M. D. Silver, and S. Nakanishi, "Upsampling with Shift Method for Windmill Correction," in *IEEE Nuclear Science Symposium Conference Record*, vol. 4, 2006, pp. 2293–2295.

[7] K. Brown and S. Zabic, "Method for reducing windmill articacts in multi-slice CT images," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 7961, 2011, pp. 79 611P–1.

[8] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP–a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in Medicine and Biology*, vol. 49, no. 11, pp. 2209–2218, 2004.

[9] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Transactions on Medical Imaging*, vol. 1, no. 3, pp. 192–196, 1982.

[10] J. Sunnegårdh and P.-E. Danielsson, "Regularized iterative weighted filtered backprojection for helical cone-beam CT," *Medical Physics*, vol. 35, no. 9, pp. 4173–4185, 2008. [Online]. Available: http://link.aip.org/link/?MPH/35/4173/1

[11] H. Turbell, "Cone-Beam reconstruction using filtered backprojection," PhD thesis 672, Department of Electrical Engineering, Linköping University, Linköping, Sweden, February 2001.

[12] G. Lauritsch and H. Bruder, "Head phantom," As available electronically in May 2009 at http://www.imp.uni-erlangen.de/phantoms.

[13] L. A. Shepp and B. F. Logan, "Reconstructing interior head tissue from x-ray transmissions," *IEEE Transactions on Nuclear Science*, vol. 21, no. 1, pp. 228–236, 1974.

Joint Estimation of Deformation and Penalized-Likelihood CT Reconstruction Using Previously Acquired Images

Hao Dang, Adam S. Wang, Zhe Zhao, Marc S. Sussman, Jeffrey H. Siewerdsen, J. Webster Stayman*

***Abstract*—Patients often undergo repeated x-ray CT acquisitions in both diagnostic imaging (e.g., lung nodule surveillance) and image-guided interventions (e.g., biopsy needle guidance). Radiation dose is a particular concern in such sequential CT studies. Traditional imaging paradigms treat each acquisition in isolation, neglecting a wealth of patient-specific anatomical information from previous studies and failing to seize an opportunity for dose reduction and/or image quality improvement. We propose a reconstruction framework that incorporates a previously unregistered patient-specific prior image as part of a penalized-likelihood (PL) reconstruction. The method jointly estimates patient motion between the original acquisition and the current data, and the image attenuation parameters that are part of traditional reconstruction. Central to this approach is a deformable registration scheme that can accommodate motion – in the case of thoracic imaging, arising primarily from respiration and variations in patient setup. To investigate the performance of this approach, we performed cadaver experiments on a cone-beam CT test-bench simulating a lung nodule surveillance scenario with highly downsampled acquisitions (a factor of 18 exposure reduction). The proposed approach yields reconstructions with a major increase in image quality compared to traditional approaches as well as model-based approaches that neglect prior information or use an unregistered prior image, suggesting great potential for dose reduction while preserving image quality.**

***Index Terms*—Deformable Motion Estimation, Prior Image Reconstruction, Penalized-likelihood Reconstruction**

I. Introduction

Sequential CT acquisitions are conducted in many clinical scenarios. In diagnostic imaging, for example, lung nodule surveillance scans are used to assess tumor growth rates through estimation of doubling times or to monitor patients following therapy. In image-guided interventions, a pre-operative scan is typically used for surgical planning, and (perhaps multiple) intra-operative scans may be acquired for up-to-date visualization of tissue change or localization of surgical tools, e.g., guidance of a needle or other interventional device. Such imaging studies are traditionally formed through a series of *complete* acquisitions, making radiation dose a particular concern. Simply reducing the exposure per projection will increase noise in reconstructions, while reducing the number of projections makes the reconstruction problem more ill-posed, often resulting in significant artifacts and thereby reducing image quality. Thus, reconstruction methods that seek to preserve image quality in low-dose imaging should better utilize noisy measurements and/or incorporate additional information to overcome limited data.

*The authors are with Johns Hopkins University, Baltimore, MD 21212 USA (corresponding author is J. W. Stayman, phone: 410-955-1314; fax: 410-955-1115; e-mail: web.stayman@jhu.edu).

This work was supported in part by Varian Medical Systems, Inc.

Model-based iterative reconstruction (e.g., penalized-likelihood estimation) is one strategy that improves the utilization of noisy measurements. The general framework of the model-based approach permits incorporation of many aspects of the physical detection process (e.g., noise, arbitrary geometry, scatter, etc.). Moreover, the framework allows for various penalties for enforcing or encouraging desirable image properties to provide improved trade-offs between dose and image quality. However, such techniques rarely leverage the rich information found in sequential studies.

Despite the wealth of information that is shared in repeated acquisitions, traditional imaging paradigms treat each acquisition in isolation, neglecting previously measured patient-specific anatomical information (e.g., prior images) from prior acquisitions and an opportunity for dose reduction. The idea of incorporating prior images directly into reconstruction has been established in methods such as Prior Image Constrained Compressed Sensing (PICCS) techniques [1-2] in which an objective encouraging similarity with the prior image is posed, but sparse differences are allowed through the use of l_1 norms. Data consistency is typically enforced through a linearized equality constraint, which may place limits on the complexity of the forward model and disregard noise. More recent PICCS techniques additionally incorporate a noise model [3].

We previously introduced a prior-image-based reconstruction approach [4] that combines both the likelihood-based framework and penalty functions that integrate prior-information. The approach allows for a great deal of flexibility in the sophistication of the forward model and noise model, since an explicit linearization of the model is not required.

A critical aspect of efficient use of previous anatomical information in any prior-image-based approach is the compensation of patient motion between acquisitions. Without motion compensation, prior-image-based reconstruction approaches cannot differentiate between *true anatomical change* (e.g., tumor growth or tissue resection) and *change due to motion.* Ambiguity between these two types of change makes true anatomical change difficult to recognize and could easily lead to introduction of false anatomical changes.

We choose to compensate patient motion by incorporating a proper 3D deformation model into a model-based iterative reconstruction method, and to use an optimization framework that jointly estimates deformation and the image reconstruction. The concept of joint estimation has been studied to recover periodic motion (e.g., cardiac and lung) from a single gated acquisition [5-7], but has had less attention within the context of prior-image-based reconstruction. In this work, we propose a framework that incorporates a patient-

specific prior image with a cubic B-spine based deformation model into penalized-likelihood estimation. An alternating maximization optimization was applied to solve the joint estimation problem, and the approach was evaluated in cadaver experiments emulating lung nodule imaging scenarios.

II. Methods

A. Forward Model and Penalized-Likelihood Estimation

In x-ray CT, a discretized object μ can be related to mean measurements $\bar{y}$ via Beer's Law, with a forward model written in matrix-vector form as

$$\bar{y} = \mathbf{D}\{g\}\exp(-l), l = \mathbf{A}\mu \tag{1}$$

where $\mathbf{D}$ is an operator converting a vector of measurement-dependent gains g to a diagonal matrix, l denotes the vector of line integrals, and $\mathbf{A}$ is the $M \times N$ system matrix (i.e., linear projection operator).

While one can apply an arbitrary noise model for the measurements, we choose a Poisson model with independent measurements y_i to yield the log-likelihood function in Eq. (2):

$$\log L(y;\mu) = \sum_{i=1}^{M} h_i\left(\left[\mathbf{A}\mu\right]_i\right) = \sum_{i=1}^{M} \log\left(g_i \exp\left(-\left[\mathbf{A}\mu\right]_i\right)\right) - g_i \exp\left(-\left[\mathbf{A}\mu\right]_i\right)$$

where h_i denotes a marginal log-likelihood for each i.

A general penalized-likelihood estimation (PLE) can then be formed as the maximizer of the log-likelihood above with a general image regularizer $R(\mu)$, written as:

$$\mu = \arg\max_{\mu} \log L(y;\mu) - R(\mu), R(\mu) = \beta_R \left\|\mathbf{\Psi}_R \mu\right\|_{p_R}^{p_R} \tag{3}$$

A specific image roughness penalty has been chosen containing an image gradient operator $\mathbf{\Psi}_R$, a p-norm, and a scalar control parameter β_R.

B. Deformable Prior Image Registration, Penalized-Likelihood Estimation (dPIRPLE)

We modify (3) with an additional prior penalty term that encourages similarity between the current estimate and the prior image μ_P after applying a deformation $\mathbf{W}(\lambda)$. We denote the proposed approach as dPIRPLE, and its objective function Φ can be written as Eq. (4):

$$\left\{\hat{y}, \hat{\lambda}\right\} = \arg\max_{\mu,\lambda} \log L(y;\mu) - \beta_P \left\|\mathbf{\Psi}_P\left(\mu - \mathbf{W}(\lambda)\mu_P\right)\right\|_{p_P}^{p_P} - \beta_R \left\|\mathbf{\Psi}_R \mu\right\|_{p_R}^{p_R}$$

where the additional prior penalty term contains its own image gradient operator $\mathbf{\Psi}_P$, a p-norm, and a scalar β_P. The p-norm value p_P can be freely chosen to achieve different performance. For example, p_P=2 tends to enforce smooth differences thereby blending features in the prior image and current measurements, while p_P=1 tends to encourage sparse differences (as one finds in compressed sensing).

The deformation $\mathbf{W}(\lambda)$ permits any suitable model. In this work, we adopted a cubic B-spline deformation model [8]:

$$\mathbf{W}_\lambda(x) = x + \sum_{x_i \in N_x} \lambda_i \beta\left(\frac{x - x_i}{\sigma}\right) \tag{5}$$

where $\beta(\cdot)$ is the tensor product of cubic B-spline functions, x_i are the control points, σ is the control point spacing, λ_i is the B-spline coefficient vectors (i.e., control point displacements), and N_x is the set of control points within the support of the B-spline at x. If $\mathbf{W}(\lambda)$ is removed in (4), the prior image μ_P will not be motion compensated, referred to simply as Prior Image, Penalized Likelihood Estimation (PIPLE).

C. Optimization Algorithm for dPIRPLE

A modified p-norm function has been implemented in which the original p-norm in a δ-neighborhood around the origin is replaced by a quadratic function so that the new function becomes differentiable at the origin. When p=1, the modified p-norm is equivalent to the sum of Huber loss functions evaluated at each element.

Although (4) becomes differentiable with a modified p-norm, it is still generally a nonconvex function of (μ, λ) and therefore challenging to optimize. The optimization can be simplified by using an alternating maximization approach in which we maximize over μ with fixed λ and vice versa, i.e.:

$$\mu^{n+1} = \arg\max_{\mu} \Phi\left(\mu, \lambda^n\right) \tag{6}$$

$$\lambda^{n+1} = \arg\max_{\lambda} \Phi\left(\mu^{n+1}, \lambda\right) = \arg\max_{\lambda} -\beta_P \left\|\mathbf{\Psi}_P\left(\mu^{n+1} - \mathbf{W}(\lambda)\mu_P\right)\right\|_{p_P}^{p_P}$$

Image update: With fixed λ, (4) becomes a standard PLE with two penalty terms. This allows for the use of well-known optimization approaches for PLE. We choose the separable paraboloidal surrogates (SPS) approach [9], which makes the image update easily parallelizable for each voxel. Both penalty terms (for $p \geq 1$) meet the five criteria in [9] so that their surrogates can be found.

Registration update: With fixed μ, (4) reduces to the prior image penalty term (dropping two constant terms) - i.e., the difference between the current estimate and the deformed prior image. This is essentially a standard image registration problem, with a modified p-norm as a similarity metric, which can be solved efficiently using existing registration software with minor modifications. For example, for p_P=2, this is equivalent to deformable registration with a Sum of Squared Difference (SSD) similarity metric. We choose to use a Limited-memory variant of BFGS (L-BFGS) [10] to approximate the Hessian matrix in our implementation. The gradient of (4) in the registration update is:

$$\left[\nabla_\lambda \Phi(\mu,\lambda)\right]_i = \frac{\partial}{\partial \lambda_i}\Phi(\mu,\lambda) = -\left[\mathbf{\Psi}_P \dot{\mathbf{W}}_i(\lambda)\mu_P\right]^T \dot{f}\left(\mathbf{\Psi}_P\left(\mu - \mathbf{W}(\lambda)\mu_P\right)\right) \tag{7}$$

A pseudo-code representation of the optimization algorithm for dPIRPLE is in **Algorithm 1**.

Algorithm 1: dPIRPLE

Input μ^0, $\lambda^{0,0}$, Inverse Hessian $\mathbf{H}^{0,0}$
for n = 0 to max_iteration - 1
  if do registration at iteration n
    for r = 1 to R **% Registration update**
      Compute gradient using (7), update $\mathbf{H}^{n,r}$ from $\mathbf{H}^{0,0}$, do line search with Wolfe condition, update $\lambda^{n,r}$ from $\lambda^{n,r-1}$
    end for
    $\lambda^{n+1,0} = \lambda^{n,R}$
  end if
  for j = 1 to N **% Image update**
    Update μ_j^{n+1} from μ_j^n
  end for
end for
return μ and λ

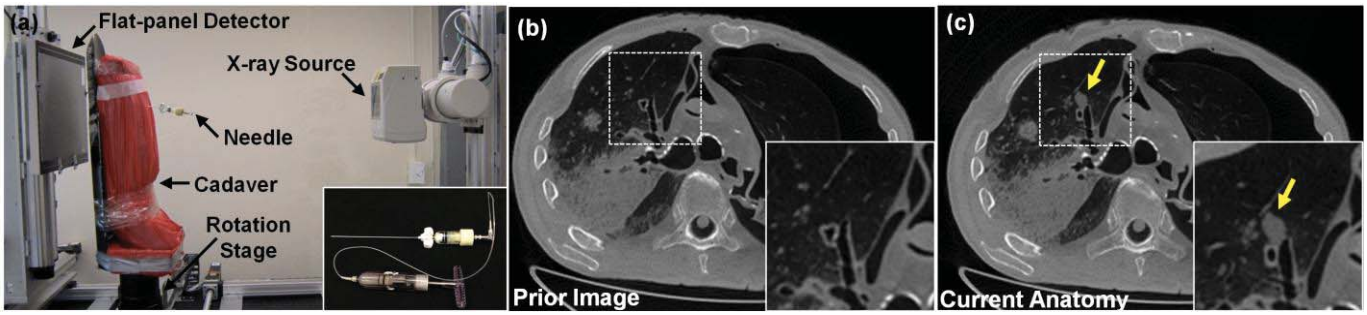


Fig. 1. (a) Cadaver and CBCT bench. (inset: petroleum jelly injector) (b) Patient-specific prior image, PLE of a complete pre-injection (nodule-absent) acquisition (360 views over 360°). (c) Current anatomy, PLE of a complete post-injection (nodule-present) acquisition. The nodule is marked by an arrow.

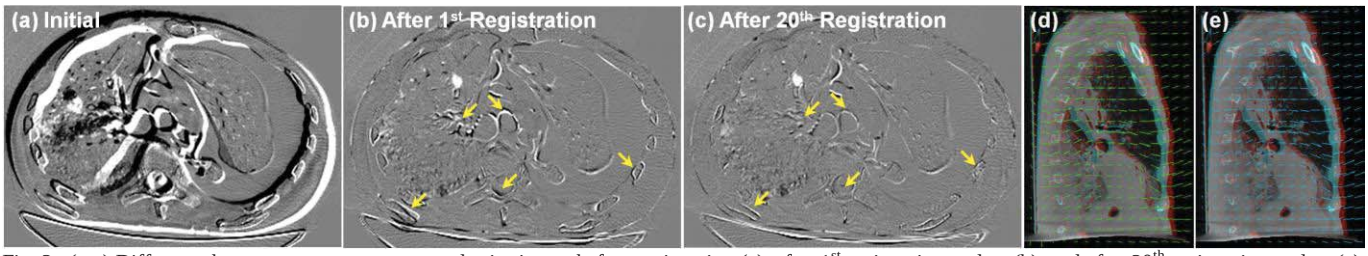


Fig. 2. (a-c) Difference between current anatomy and prior image before registration (a), after 1st registration update (b), and after 20th registration update (c). (d) Final deformation field estimated by dPIRPLE. (e) 'Optimal' deformation field acquired by two fully sampled datasets. Each vector represents the in-plane displacement of one voxel with vector magnitude scaled by a factor of 2 for visualization.

Whereas keeping the objective function strictly the same in two updates is preferred in solving a single objective function, in practice, one may choose different p_P in the two updates. This can yield desirable convergence performance in both registration (where low p values can lead to poor convergence) and the image update (which requires low p values to encourage sparse differences). We implemented the dPIRPLE algorithm in Matlab, with computationally intensive functions (e.g., projection/backprojection) calculated using CUDA-based libraries and the registration toolbox *Elastix.* [11]

D. Cadaver Experimental Methods

We conducted cadaver experiments on a flat-panel cone-beam CT (CBCT) test-bench (Fig. 1(a)) to evaluate the dPIRPLE approach. The imaging task was to reconstruct a newly formed lung nodule (introduced between scans) in the presence of deformable patient motion between two acquisitions. The detector (PaxScan 4343CB, Varian Medical Systems, Palo Alto, CA) had 1536×1536 pixels at 0.278×0.278 mm^2 pixel pitch after 2×2 binning. The system geometry involved a 150 cm source-to-detector distance and 120 cm source-to-axis distance. All data were reconstructed with 260×300×330 voxels and 1×1×1 mm^3 voxel size.

To simulate tumor growth between acquisitions, ~1 cm^3 Petroleum jelly (~0.013 mm^{-1} attenuation) was injected into the right lung of the cadaver by a thoracic surgeon. Two fully sampled datasets (360 views over 360° at 100 kVp and 450 mAs) were acquired before and after the injection. PLE (p_R=2, β_R=10^6) was applied to both the nodule-absent and nodule-present datasets to obtain the patient-specific *prior image* and *current anatomy*, respectively (Fig. 1(b-c)). Motion imparted between scans is evident in the difference image in Fig. 2(a). From the post-injection data, 20 projections equally spaced over 200° (25 mAs) were selected to simulate an undersampled low-dose follow-up lung surveillance image at 1/18 the exposure of a fully sampled acquisition.

III. Experimental Results

We reconstructed highly undersampled projections using FBP (filtered-backprojection), PLE (FBP initialization, p_R=1, β_R=10), PIPLE (PLE initialization, p_P=1, β_P=10^4, p_R=1, β_R=$10^{3.5}$), and the proposed dPIRPLE (same parameters as PIPLE but with p_P=2 in registration update). In dPIRPLE, each registration update consisted of four levels of image pyramids (10×10×10 mm^3 grid size at the finest level) followed by 50 image updates. We choose $\mathbf{\Psi}_P$ equal to identity, $\mathbf{\Psi}_R$ as the first-order spatial difference operator, and δ=10^{-4} mm^{-1} in the modified p-norm. The total run time for 50 iterations was about 4 min for PLE, 7.5 min for PIPLE, and 14.5 min for dPIRPLE on a high performance workstation.

A. Convergence

Fig. 2 (a-c) shows the evolution of the deformation estimate versus iteration number. Specifically, we show the difference between the deformed prior image and current anatomy at different stages. We observe that most motion was compensated after the 1st registration update, which substantially prevented incorrect structures from being injected into subsequent image updates. The remaining differences (arrows in Fig. 2(b-c)) continued to be reduced after 20 updates, demonstrating the importance of joint estimation. The final deformation field estimated by dPIRPLE (Fig. 2(d)) closely matches an 'optimal' deformation field derived from

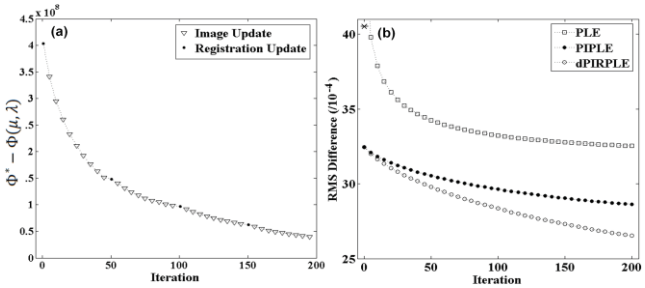


Fig. 3. (a) dPIRPLE objective function difference versus iteration. (b) Root Mean Square (RMS) difference from current anatomy versus iteration for PLE, PIPLE, and dPIRPLE. The asterisk indicates Iteration 0 outside the plot.

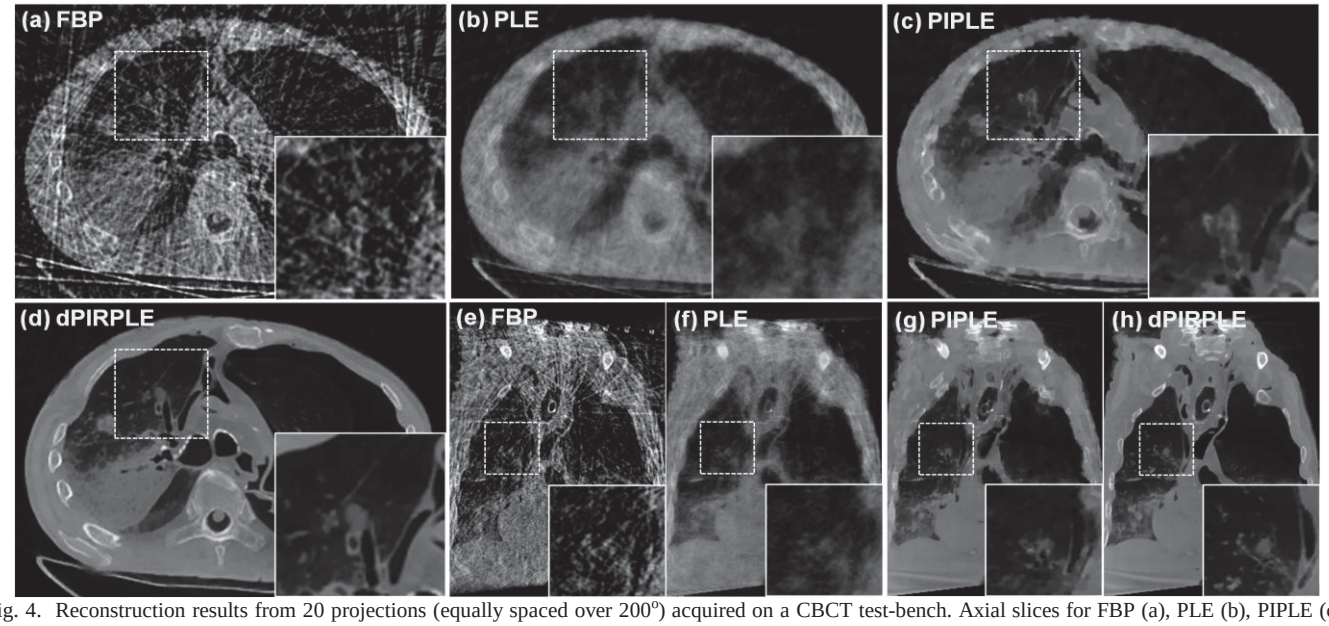


Fig. 4. Reconstruction results from 20 projections (equally spaced over 200°) acquired on a CBCT test-bench. Axial slices for FBP (a), PLE (b), PIPLE (c), dPIRPLE (d); coronal slices for FBP (e), PLE (f), PIPLE (g), dPIRPLE (h).

the fully sampled reconstructions (Fig. 2(e)). We computed root mean square (RMS) differences between the current anatomy and the deformed prior image in a Region of Interest (ROI, 100×100×100 voxels centered on the nodule): 76.5×10^{-4} mm^{-1} (at iteration zero), 24.6×10^{-4} mm^{-1} (after 20 registration updates), and 22.6×10^{-4} mm^{-1} (with 'optimal' registration).

Fig. 3(a) shows the objective function difference for dPIRPLE as a function of iteration. The objective function value at the solution, Φ^*, is estimated using 1000 iterations of dPIRPLE. The objective increases monotonically within every 50 image updates due to the monotonicity of SPS. The 1st registration yields a dramatic objective increase with smaller increases in subsequent registrations, consistent with Fig. 2. In plots of RMS difference from current anatomy versus iteration (Fig. 3(b)), PLE quickly reduced RMS difference, but plateaued. In contrast, dPIRPLE saw reduction throughout, and reduced 2× faster than PIPLE.

B. Reconstruction

Fig. 4 compares the reconstruction results from different approaches. We ran 1000 iterations of PLE, PIPLE, and dPIRPLE to generate (nearly) converged images. While FBP exhibits substantial artifacts that would defy confident nodule detection, PLE has reduced artifact but low spatial resolution owing to strong regularization and rendering makes the nodule difficult to detect. PIPLE exhibits higher contrast in the nodule area, but severe mismatches due to lack of proper registration distorts the nodule shape and introduces a considerable number of ambiguous structures which *do not reflect true anatomy*. Finally, dPIRPLE presents an accurate estimate of the true anatomy and renders the nodule clearly. The RMS difference from current anatomy within the ROI was 36.8×10^{-4} mm^{-1}, 31.1×10^{-4} mm^{-1}, and 31.9×10^{-4} mm^{-1} for FBP, PLE, and PIPLE, respectively, and 16.3×10^{-4} $mm^{-1}$for dPIRPLE.

IV. Conclusion

The proposed dPIRPLE approach demonstrates a major improvement in image quality under conditions of highly undersampled data compared to traditional approaches and model-based approaches that neglect prior information or use an unregistered prior image. This suggests that the dPIRPLE approach could be valuable in clinical scenarios offering a patient-specific prior and requiring dose reduction without loss in image quality. The dPIRPLE approach also shows that the joint maximization estimates patient motion more accurately than staged registration and yields an improved representation of true anatomy without false structures arising from misregistration. A limitation of the current work includes a residual registration error, which we plan to reduce using more sophisticated registration techniques, such as [12].

References

[1] G.H. Chen, *et al.*, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.,* vol. 35, pp. 660–663, 2008.

[2] B. Nett, *et al.*, "Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (PICCS) on a C-arm system," *SPIE Med. Imag.*, pp. 6913, 2008.

[3] P. T. Lauzier, and G. H. Chen, "Characterization of statistical prior image constrained compressed sensing (PICCS): II. Application to dose reduction," *Med. Phys.,* vol. 40, 2013.

[4] J. W. Stayman, *et al.*, "Penalized-likelihood reconstruction for sparse data acquisitions with unregistered prior images and compressed sensing penalties." *SPIE Med. Imag.*, pp.79611L-1-79611L-6, 2011.

[5] D. R. Gilland, *et al.*, "Simultaneous reconstruction and motion estimation for gated cardiac ECT," *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 2344-9, 2002.

[6] J. A. Fessler, "Optimization transfer approach to joint registration / reconstruction for motion-compensated image reconstruction." *Proc. IEEE Int. Symp. Biomed. Imaging*, pp. 596-599, 2010.

[7] K. Taguchi, *et al.*, "Image-domain motion compensated time resolved 4D cardiac CT," *SPIE Med. Imag.*, pp. 651016, Mar. 2007.

[8] D. Rueckert, *et al.*, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, pp. 712-721, 1999.

[9] H. Erdogan, and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, pp. 801-814, 1999.

[10] D. C. Liu, and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Prog.*, vol. 45, pp. 503-528, 1989.

[11] S. Klein, *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.,* vol. 29, pp. 196-205, 2010.

[12] A. Uneri, *et al.,* "Deformable registration of the inflated and deflated lung in cone-beam CT-guided thoracic surgery: Initial investigation of a combined model-and image-driven approach," *Med. Phys.,* vol. 40, 2013.

Sparse Sampling for CT Dose Reduction

Kevin M. Brown[1], Thomas Koehler[2] , Frank Bergner[2], Rolf Bippus[2], Bernhard Brendel[2], Stanislav Žabić[1], W. Clem Karl[3], Sarabjeet Singh[3], Atul Padole[3], and Synho Do[3].

***Abstract*—Two approaches are compared for reducing dose in CT imaging, both using a penalized-likelihood algorithm for reconstruction: reducing dose by reducing the tube current, and reducing dose by acquiring fewer projections (sparse sampling). Results show that the sparse sampling approach has advantages at dose levels yielding significant photon starvation, and the advantages grow as the dose becomes smaller.**

***Index Terms*—CT, reconstruction.**

I. Introduction

The concept of tomographic reconstruction from sparsely sampled data (also known as compressed sensing) has been an area of much research in recent years. In such acquisitions, projections are acquired in a non-continuous fashion, often with large angular gaps between subsequent projections. Sparse sampling schemes have been investigated in detail in many tomographic imaging modalities, including breast CT [1], small animal imaging [2], rotational x-ray [3], and radiotherapy planning [4]. One advantage of a sparse sampling approach in these modalities is that often a significant reduction in the time to acquire the scan can be achieved.

In medical CT however, up to this point there has been somewhat less motivation for adopting sparse sampling strategies. This is partly due to the fact that no rotation speed or scan speed advantage is gained, since in most medical CT scanners the rotation speed is limited mainly by the gantry hardware, and not by the detector readout speed or tube power. In other words, acquiring fewer projections does not allow one to scan any faster. In addition, most currently used x-ray tubes in medical CT scanners are designed for continuous high-flux utiliztion, and do not support the rapid on and off switching of the x-ray tube current required to support these sparse sampling strategies. As a result, only very little work (viz., [5]) has been published in the area of clinical application of sparse sampling in CT so far.

However, the desire to significantly lower doses for routine CT scans raises a motivation to investigate sparse sampling for clinical CT. For very low tube currents or large patients, very few photons will penetrate to the detector, and many detector pixels will measure no photons during an integration period. This means that such a projection will contribute very little real information into the resulting reconstruction. In addition, at severely low photon levels, effects from detector electronic noise and non-linearities become more pronounced, and can lead to undesireable image artifacts.

[1] Philips Healthcare, Cleveland, OH. E-mail: kevin.m.brown@philips.com.
[2] Philips Technologie GmbH, Innovative Technologies, Research Laboratories, Hamburg, Germany
[3] Massachusetts General Hospital, Boston, MA

One strategy to overcome the photon starvation problem is to acquire fewer projections, and increase the dose in each projection. This can be done easily by extending the integration period of each projection, and still acquiring in a continuous fashion. However, this approach comes with a corresponding loss of spatial resolution, due to the additional angular blur in each projection. Acquiring fewer projections in a sparse fashion might overcome the photon starvation problem with less penalty in terms of resolution, thus enabling scans at significantly lower doses than can be achieved today.

An interesting question then becomes: given a tube with the ability to deliver dose either in a low-flux, continuous fashion, or a high-flux, sparse fashion, which approach yields the best image quality for a fixed dose? Jørgensen et. al. [1] made an attempt to evaluate this trade-off in the domain of breast CT, and conclude that higher sampling rates always yield the best image quality. However, the problem of photon starvation was not specifically addressed in this work, and the image quality was judged only for simulated breast micro-calcifications. Han et. al. [4] performed a similar study for CBCT, but again dose levels which yield photon-starved projections were not considered. In this paper we investigate sparse sampling schemes for medical CT with the goal of achieving sub-mSv dose levels for large BMI patients.

The inclusion of clinical data presents an additional problem: how to generate different datasets of the same patient for various sparse and low-dose sampling schemes, without performing additional patient scans. Fortunately, Philips has been developing software which can accurately simulate low-dose projections from original high-dose scans [6], including accurate models of the electronic noise. Sparse scans can be simulated simply by selecting subsets of projections from the original scan, or from simulated lower mA datasets. Thus, a complete set of sparse sampling schemes can be generated from a single routine-dose patient dataset, eliminating problems such as patient motion or changing contrast levels, which might arise when comparing multiple real scans.

The issue of reconstruction algorithms also comes into play, since traditional FBP reconstruction suffers from aliasing and streaking artifacts when applied to sparse data. The plethora of papers on this topic in recent conferences (see for example Jørgensen [1], Langet [3], and Han [4], as well as Sidky [7] and Huang [8]) shows that the optimal reconstruction algorithm for sparsely sampled data is by no means a resolved problem. In this work we evaluate a penalized maximum-likelihood algorithm on both continuous and sparsely-sampled acquisitions, at a constant low dose.

II. Methods

A. Data Generation

We chose a clinical CT dataset (Philips iCT) with a reasonably large patient (108 kg), such that an aggressive dose reduction would likely lead to severe photon starvation. The original patient scan was at 150 mAs (200 mA), resulting in an effective dose of 7.4 mSv. Low-dose datasets where then generated from this original data in one of two ways.

First, selecting a subset of the projections from the original dataset, such as every 8th projection, yields a low-dose dataset with a uniform sparse angular sampling pattern. We abbrevate this method of dose reduction as a sparse sampling factor (SSF).

The second method involves regenerating projections from the original dataset to simulate projections acquired at a lower tube current and dose. The low-dose simulation follows the method given in [6], and accurately accounts for both quantum and electronic noise, as well as taking into account the inherent noise in the original dataset. In this way, problems occuring from multiple scans (such as the movement of the patient) can be eliminated while still allowing the use of clinical data for the evaluation. We abbreviate this method of dose reduction as a tube current reduction factor (TCF).

A total dose reduction can then be achieved either independently by TCF or SSF, or by a combination of the two (the product of TCF and SSF gives the total dose reduction factor). For this study we compared 3 strategies of progressive dose reduction which yield the same total dose:

1) SSF=1 (full sampling), with TCF (simulated tube current) of 8, 10, 12, 15, 20 and 30.
2) TCF=1 (full tube current), and SSF (sparse sampling) of 8, 10, 12, 15, 20 and 30. This is also denoted as sampling strategy $s1$.
3) TCF=2 (half tube current), and SSF of 4, 5, 6, 7.5, 10 and 15 (sampling strategy $s2$).

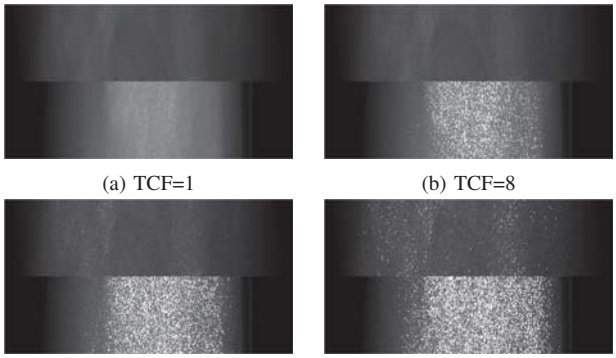
To give one example, a dose reduction factor of 20 can be achieved by SSF=1 and TCF=20, or SSF=20 and TCF=1, or SSF=10 and TCF=2. Effective doses thus range from 0.9 mSv down to 0.25 mSv for the cases listed above.

An important point is that only method (1) is achievable on most clinical CT systems today, and there is often a limit to the minimum tube current as well. Furthermore, Figure 1 shows how the projection quality in the ventral and lateral directions is degraded at very low tube currents. White spots indicate detectors which measured no photons during a given integration period (or where electronic noise caused the signal to go negative). At a TCF=30, these points represent about 20% of the detector pixels measuring line integrals through the patient in the lateral projection. The primary motivation then in this study is to investigate the sparse sampling approach as a means for aggressively lowering dose while avoiding these photon starvation effects.

B. Reconstruction Algorithm

For reconstruction we use a penalized maximum-likelihood approach, with a cost function of the form

$$\Psi(u) = -L(Au|y) + \beta R(u) \qquad (1)$$



(a) TCF=1 (b) TCF=8

(c) TCF=15 (d) TCF=30

Fig. 1. Example log projections along the ventral (top) and lateral (bottom) axes, for varying tube current factors (TCF). White spots indicate detectors that measured no photons during the given projection, or where the signal became negative due to electronic noise.

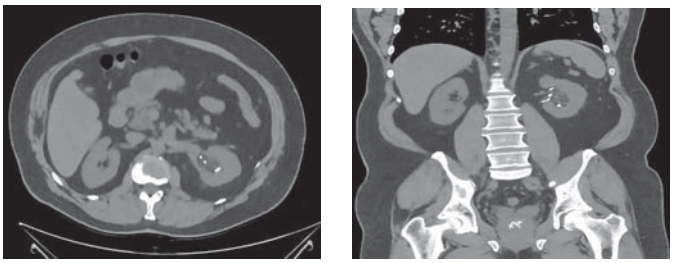


Fig. 2. Reference images ($\delta = 2$ HU, $\sigma = 3$ HU). W/L = 500/50 HU.

with system matrix (or forward projection operation) A, an image u, measured data y, and roughness penalty $R(u)$. As an algorithm to solve for an image which optimizes this cost function, we take the ordered-subsets version of the separable paraboloidal surrogates (SPS) algorithm [9]. For the roughness penalty $R(u)$ we use the Huber function with the transition from quadratic to linear at a value of δ. For the regularization strength β the algorithm uses an automated regularization controller as described in [10] to achieve a pre-selected target noise level σ (pixel standard deviation) in the image.

We reconstructed images with the following pairs of parameters:

- $\delta = 2$, 5, 10, and 20 HU
- $\sigma = 5$, 7, 7, and 10 HU

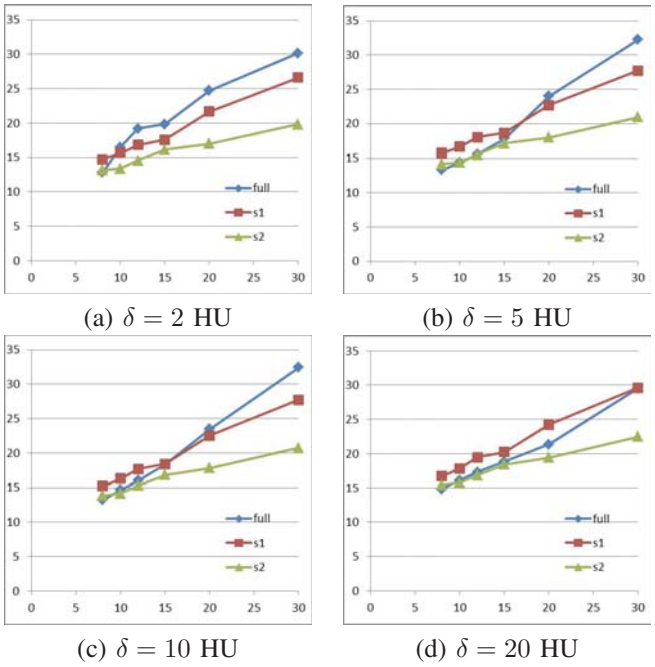
The likelihood function assumes Gaussian noise for the higher tube currents (TCF=1 and 2), and Poisson noise for TCF≥ 8[1]. Note that other than the likelihood and the change in the data size, there is no difference in the algorithm used for the fully sampled (SSF=1) or sparsely sampled (SSF> 1) datasets, and that our β-selection method ensures that we compare images with the same noise for a given value of δ.

III. Results

Figure 3 shows plots of the Mean Absolute Error (MAE) [2] of the dose-reduced images relative to reference images

[1]The Gaussian noise model was also investigated for the low tube currents, and found to yield images with a larger bias than the Poisson model.

[2]We evaluated also the root mean square metric with qualitatively the same results.



(a) $\delta = 2$ HU (b) $\delta = 5$ HU

(c) $\delta = 10$ HU (d) $\delta = 20$ HU

Fig. 3. Mean Absolute Error (in HU) from the original full-dose images, calculated over the whole volume. x-axis is the total dose reduction factor

reconstructed from the original full dose dataset (see Figure 2), for different values of δ. The mean is taken over the entire 3D image volume. While MAE is not a fool-proof method of evaluating image quality, some general trends are clear from the graphs. The first is that the MAE is always lower for sampling s2 than for s1. This agrees with the result from [1] that higher sampling rates are generally better, when the flux is high enough to avoid photon starvation. Figure 4 shows one comparison between the two sampling rates for a dose reduction factor of 15; the angular aliasing artifacts can be clearly seen with SSF=15.

The second important observation is that, while methods full-sampling and s2 have about the same MAE at a dose reduction factor of 8, as the dose is lowered further the error in the fully-sampled images increases much faster than the error in the sparsely-sampled images. This points to an advantage of the sparse sampling at extremely low doses, where this approach allows avoiding severely photon-starved projections.

A third point is that the MAE values are very similar for different values of δ. This is somewhat surprising given the preference of many researchers for the total variation (TV) penalty in sparse sampling applications (which is most similar to the $\delta = 2$ case), but the graphs suggest that the Huber penalty also offers a flexible way to adjust image quality by allowing images with less "flat" noise patterns than TV.

The remaining figures show some example images which give a visual confirmation of the results discussed above. Figure 5 shows axial images comparing TCF=2 and SSF=1 at equivalent doses, for the algorithm setting $\delta = 2$ HU. Note that three calcifications in the kidneys are visible at all dose levels for the sparse sampling datasets (5-a), but one calcification rapidly disappears in in the fully-sampled images (5-b, red arrow). Also the ribs rapidly become blurry in the

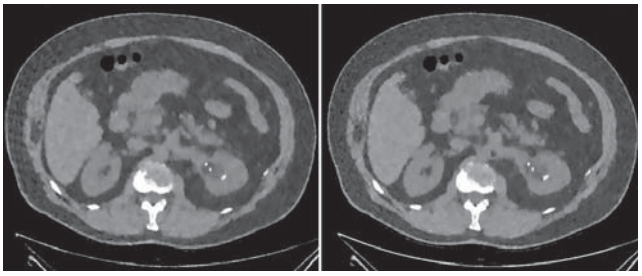


Fig. 4. Axial images comparing SSF=15 (left) and SSF=7.5 (right) for an equal dose reduction factor of 15.

fully-sampled images. Figure 6 shows the same comparison for a coronal slice. There a change in the resolution of the objects in the kidneys can be observed as well for full-sampling, but more apparent is the degradation in the pelvis region (red oval) for the fully-sampled images.

Figure 7 shows axial and coronal images for the setting $\delta = 10$ HU, for a total dose reduction factor of 20, with sampling $s2$ and full sampling. A reasonable image quality is maintained here as well using $s2$, with a less "flat" noise pattern which could be more acceptable to radiologists.

IV. Conclusion

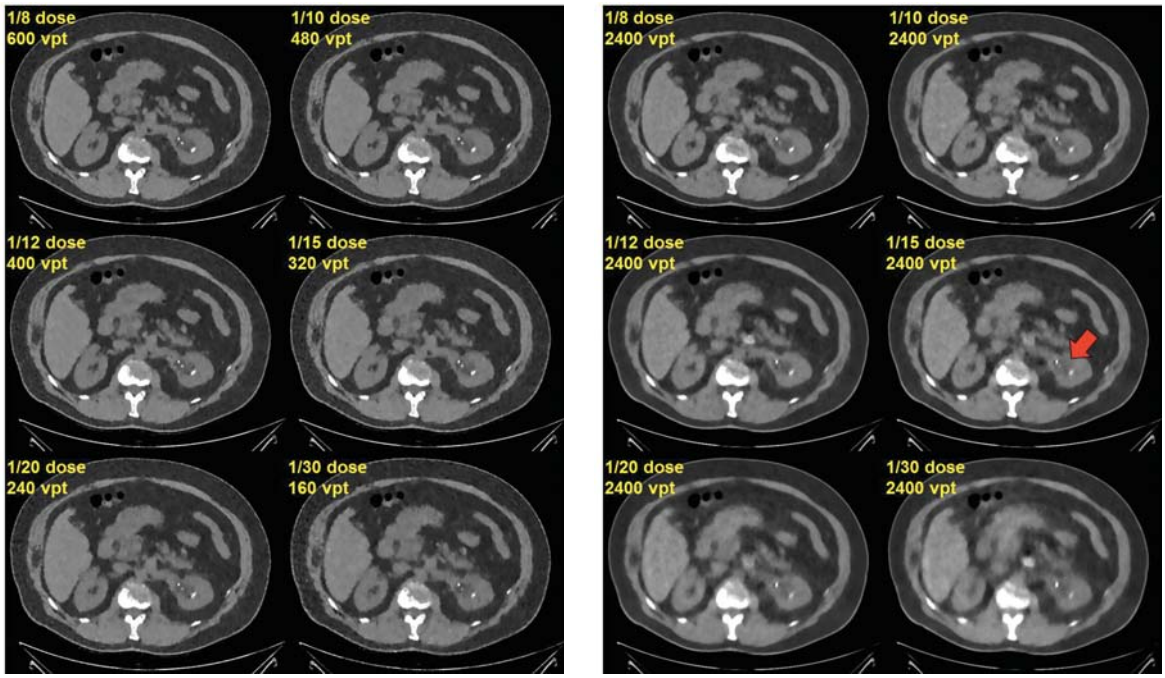
We have demonstrated that sparse sampling can deliver improved image quality over tube-current reduction for aggressive low-dose scans. In particular, the value of sparse sampling increases as the total dose is lowered, or more specifically, as the photon-starvation problem increases. In other words, the results show that image quality always decreases as the total dose decreases, but it decreases at a slower rate if sparse sampling is used. Future work may involve determining minimum dose levels and optimal sampling patterns for a range of clinical applications.

V. Acknowledgements

The authors gratefully acknowledge Dr. Mannudeep Kalra for review of the images and helpful discussion.

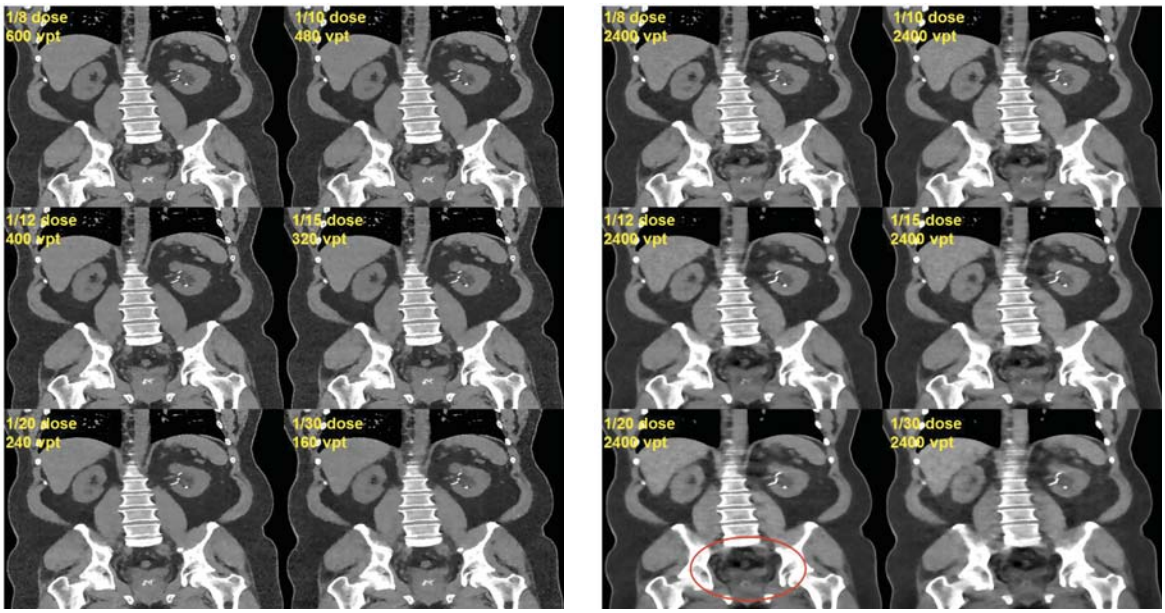
References

[1] J. H. Jørgensen et al. *Proc. Fully 3D*, (2011) 359.
[2] B. Vandeghinste et al. *Proc. Fully 3D*, (2011) 431.
[3] H. Langet et al. *Proc. Fully 3D*, (2011) 427.
[4] X. Han et al. *Proc. IEEE MIC*, (2010) 2051.
[5] G. Chen et al. *Med. Phys.*, 35 (2008) 660.
[6] S. Žabić et al. *Med. Phys.*, 40(3) (2013) .
[7] E. Y. Sidky et al. *Phys. Med. Biol.*, 57(10) (2012) 3065.
[8] J. Huang et al. *Computers in Biol. and Med.*, 41(4) (2011) 195.
[9] H. Erdogan and J. Fessler. *Phys. Med. Biol.*, 44(11) (1999) 2835.
[10] F. Bergner et al. *Proc. CT Meeting*, (2012) 267.



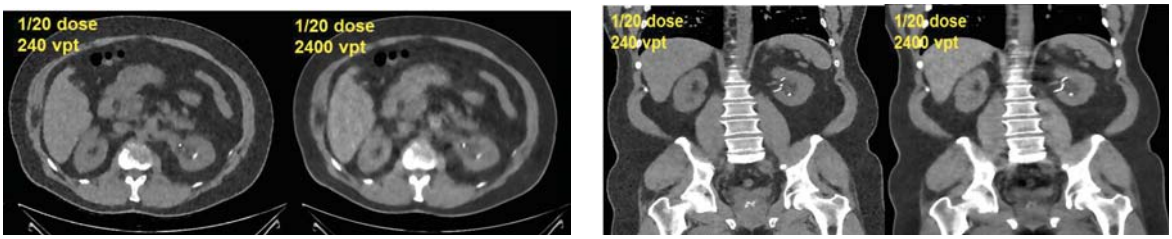
(a) TCF=2, and various SSF (method $s2$). (b) SSF=1 and various TCF (full sampling).

Fig. 5. Axial example images for $\delta = 2$ HU, $\sigma = 5$ HU. W / L = 500 / 50 HU. vpt = number of views per turn of the x-ray tube.



(a) TCF=2, and various SSF (method $s2$). (b) SSF=1 and various TCF (full sampling).

Fig. 6. Coronal example images for $\delta = 2$, HU $\sigma = 5$ HU. W / L = 500 / 50 HU. vpt = number of views per turn of the x-ray tube.



(a) TCF=2, and SSF=10 (left) and TCF=20, SSF=1 (right). (b) TCF=2, and SSF=10 (left) and TCF=20, SSF=1 (right).

Fig. 7. Axial (a) and coronal (b) example images for $\delta = 10$ HU, $\sigma = 7$ HU. W / L = 500 / 50 HU.

Challenges posed by statistical weights and data redundancies in iterative X-ray CT reconstruction

K. Schmitt, H. Schöndube, K. Stierstorfer, J. Hornegger, F. Noo

***Abstract*—Statistical iterative reconstruction methods are currently under extensive investigation for x-ray computed tomography. Among many options, the maximum likelihood solution is often preferred, particularly because it can be reduced to a weighted least-square problem. This solution may be sought using a moderate number of iterations of a specific algorithm such as Landweber's method (or an ordered-subset variant of this algorithm), or may be sought with a penalty term and a number of iterations large enough to reach convergence. In the first case, the iteration number serves as a regularization means, whereas in the second case the penalty term defines the regularization procedure. It is well-known that the iteration number creates a (shift-variant) trade-off between resolution and noise, and that such a trade-off has been found useful in nuclear medicine. In this work, we show that the noise-resolution trade-off introduced by the iteration number is not always attractive for CT imaging, particularly when statistical weights and data redundancies are involved.**

I. Introduction

Statistical iterative reconstruction methods are currently under extensive investigation for x-ray computed tomography (CT), as they may offer significant gains in terms of image quality at equal dose, and thereby allow reduction in x-ray dose. There are many ways to formulate a statistical reconstruction method for x-ray CT. In particular, the maximum likelihood solution without and with constraints on the image appear both highly popular, particularly because finding this solution in CT can be reduced to a weighted least-square problem. In the first approach, the user formulates an iterative algorithm that converges towards the maximum likelihood solution and defines the reconstruction as the application of a finite number of iteration steps. Using this approach, the iteration number is essentially seen as a regularization means. Recall that the CT reconstruction problem is mildly ill-posed, so that regularization is essential to achieve satisfactory image quality. In the second approach, the regularization is not left to the iteration number; it is enforced directly by the constraint and the user iterates as long as needed to reach to minimum of the objective function. Popular constraints include the generalized Gaussian prior and the Huber penalty.

In this paper, we investigate the usefulness of the first approach for CT imaging, that is we study the effectiveness of regularization based on a finite number of iterations. Our study includes essential aspects of CT imaging, namely non-uniform statistical weights and data redundancies.

K. Schmitt, H. Schöndube and K. Stierstorfer are with Siemens AG, Healthcare Sector. J. Hornegger is with the Pattern Recognition Lab, University of Erlangen-Nürnberg, Erlangen, Germany. F. Noo is with the Department of Radiology, University of Utah, Salt Lake City, Utah, USA.

The concepts presented in this paper are based on research and are not commercially available. This work was partially supported by NIH grant R01 EB007236; its content is solely the responsibility of the authors and do not necesarily represent the official views of the NIH.

II. Experimental setting

A. Data simulation

All simulations were performed in fan-beam geometry (3rd generation CT curved detector) using the FORBILD head phantom. Thus, each ray was parameterized by two angles, λ and γ, where λ is the polar angle specifying the source position and γ is the angle between the ray and the line that connects the source to the rotation center.

Full scans and short scans were both considered with parameters given in table 1. Also, note that a sub-sampling of each detector was employed to model the blurring that results from the finite size of the detector elements, and thereby mitigate high-frequency errors in the reconstruction. Specifically, each measurement was simulated as the average of five line integrals equally spaced over the detector width (non-linearities were neglected).

B. Image representation and forward projection model

We decided to represent the attenuation function by its values on a Cartesian grid of points. The coordinates associated with this grid are called x and y. 350 by 350 locations were considered with a uniform sampling distance of 0.075 cm in both x and y.

The link between the attenuation function and the measurements was described using the principles of the distance-driven method [1]. That is, each line integral was evaluated as a simple sum in x or y together with a linear interpolation between grid points in y or x, respectively. As suggested by this method, note that the direction of summation was fixed for all lines within a fan-beam view, i.e., the position of the x-ray source defined the summation direction for all rays within the view. Also, the interpolation kernel accounted for both the sampling distance in γ and the sampling distance in x (or y depending on the interpolation direction).

C. Reconstruction technique

Let c be the vector of unknown image coefficients, let g be the vector grouping the CT measurements, and let A be the matrix that links c to the CT measurements, as defined by the distance-driven method. Each reconstruction was performed using a moderate number, m, of iterations given by the following equation:

$$c^{(n+1)} = c^{(n)} + \eta \cdot A^T C^{-1} \left(g - A c^{(n)} \right) \qquad (1)$$

where C is a constant diagonal matrix and η is a factor controlling convergence speed. The value for η was chosen as 0.90 times $2/\sigma_{max}$ where σ_{max} is the maximum singular value of the matrix $C^{-1/2}A$, estimated using three iterations of the Power method [2]. The initial image vector, $c^{(0)}$ was always chosen as the $zero$ vector.

Conceptually, our iterative procedure can be interpreted as the application of Landweber's method to find the minimum-norm minimizer of

$$J(c) = ||C^{-1/2}(Ac - g)|| \,. \tag{2}$$

However, this minimizer was never reached since reconstruction was based on a moderate number of iterations. Recall that singular value decomposition analysis reveals that using a finite number of Landweber's iterations amounts to performing minimization with regularization. Given that the CT reconstruction problem is mildly ill-posed, regularization is actually essential. That is, the minimum-norm minimizer of $J(c)$ is not attractive.

It is well-known that the iteration number generates a trade-off between resolution and noise. In PET imaging, this trade-off is often used to select the number of iterations. In this work, we will show that the resolution-noise trade-off induced by the number of iterations is unfortunately not often attractive for CT imaging, because it does not account for reconstruction errors other than resolution effects.

III. Reconstruction with statistical weights

In this section, matrix C is interpreted as the covariance matrix for the CT measurements. Because these measurements are assumed to be statistically independed, C is a diagonal matrix, with each element on the diagonal representing the variance of one measurement. This variance is equal to the inverse of the number of photons reaching the detector, which itself is influenced by the incoming number of photons, the shape of the bowtie filter and the attenuation property of the interrogated object.

A. Bowtie filter model

On x-ray CT scanners, the bowtie filter is a shaped piece of material (usually metal) which is placed between the x-ray source and the patient. It is designed to equalize the intensities of the rays hitting the detector for a given attenuating object. The primary purpose of the bowtie filter is to decrease the patient radiation dose near the edges of the scanning field-of-view (FOV) [3].

The effect of the bowtie filter is to make the number of photons going into the scanned object, N_{in}, vary with γ. In presence of a bowtie filter, we model the number of photons entering the interrogated object as

$$N_{in}(\gamma) = N_0 \cdot e^{-\eta(\gamma)} \,, \tag{3}$$

where N_0 is the number of photons leaving the source and $\eta(\gamma)$ is a function which models the effect of the bowtie filter.

Let μ_{BF} be the linear attenuation coefficient for the bowtie filter and let d_{BF} be its thickness at $\gamma = 0$. Then η is defined as

$$\eta(\gamma) = g_{ob}(0) - g_{ob}(\gamma) + d_{BF} \cdot \mu_{BF} \,, \tag{4}$$

Fanbeam scanning parameters:	
Source trajectory radius R	57 cm
Source to detector distance (D)	104 cm
Number of detector elements	340
Angular detector width ($\Delta\gamma$)	0.1368 radians
full scan:	
Number of projections per turn	2320
short scan:	
Number of projections per turn	870
short-scan start angle λ_s	0 radians
short-scan stop angle λ_e	$3\pi/4$ radians
angular interval d	0, 5, 30
Bowtie filter:	
Bowtie filter radius (r_{BF})	8.0 cm
Bowtie filter attenuation (μ_{BF})	0.54 cm^{-1}
Bowtie filter thickness parameter (d_{BF})	0.50 cm

TABLE I

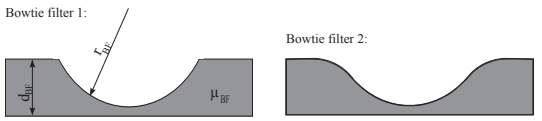


Fig. 1. Some expamples for the shape of a bowtie filter.

where g_{ob} is the line integral through the object along the line $L(\lambda, \gamma)$ for some fixed λ. Note that if the object is centered at the origin, $\eta(\gamma)$ is circularly symmetric, and any fixed value of λ will suffice for the definition of g_{ob}.

B. Experiment details

A full scan acquisition with a bowtie filter is used. We consider two phantoms: the FORBILD head phantom and a simpler phantom that consists only of the outer two ellipses of the FORBILD head phantom. Figure 1 show two bowtie filters. Here, we used a bowtie filter shaped as bowtie filter 1 and 2, respectively. The value of each statistical matrix element is defined by $C_{ij}^{-1} = N_{in} \cdot \exp(-\mu_W \cdot g_{m_{ij}})$, where N_{in} are the remaining photons after the bowtie filter (Eqn. 3) to go through the phantom, μ_W is the attenuation factor of water and $g_{m_{ij}}$ is an element of the fan-beam data set. Bias was evaluated as the mean reconstruction error over pixels located within the dark ring in Fig. 5. The error for any given pixel was defined as the absolute difference between the reconstructed value and the true attenuation value of the hatched area, which is 50 HU.

C. Results

Figure 2 shows the reconstructed images for the two phantoms with bowtie filter 1 and bowtie filter 2, respectively. The number of the photons after bowtie filter 1 and bowtie filter 2 is shown in figure 3. Figure 4 shows the bias as a function of the iteration number for bowtie filter 1 and bowtie filter 2.
Using bowtie filter 1 in our reconstruction algorithm (Eqn. 2) cause a ring artifact in the reconstructed images. This ring artifact can be traced back to the fact that the function $N_{in}(\gamma)$ is not smooth since the radius of that ring has the same radius as the one of bowtie filter, r_{BF}. By iterating a very long time the thickness of the ring decrease. Finally, the ring disappears completely first after 1640 iterations. When we use a more complex phantom, we see similar artifacts

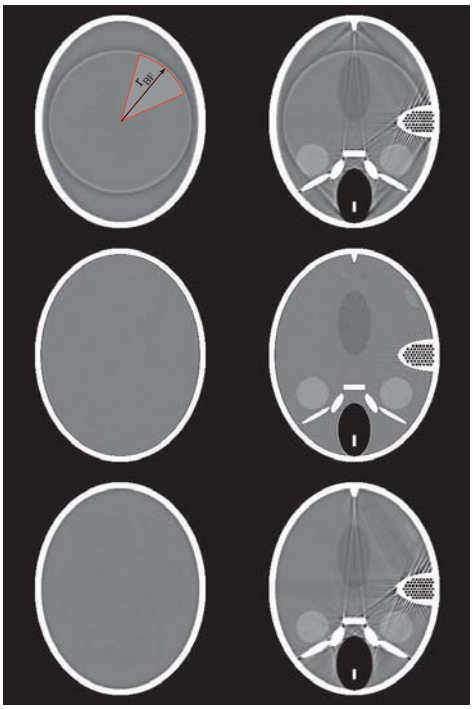


Fig. 2. Iterative reconstruction obtained from (first row) 250 iterations and (second row) 1640 iterations of the Landweber algorithm using bowtie filter 1, (third row) iterative reconstruction obtained from 250 iterations of the Landweber algorithm using bowtie filter 2. Grayscale: [1 1.1].

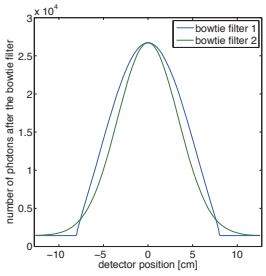


Fig. 3. Number of photons after the bowtie filter shown in Fig. 1 for different radius.

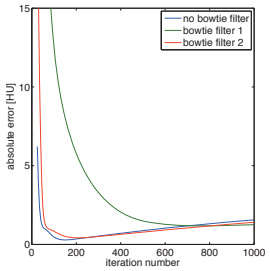


Fig. 4. Bias as a function of the iteration number by varying the number of iterations by steps of 5 beginning with 25 iterations.

appearing due to lack of smoothness in the statistical weights, but now replacement of the statistical weight with a smoother expression becomes a challenging question.

IV. Data Redundancies

It is often the case that a CT scan involves redundant data. For example, each line integral is measured twice when a full scan is performed. However, not all line integrals are always measured the same number of times. In particular, when performing a short scan, some line integrals are measured twice whereas others are only measured once. In analytical reconstruction, non-uniform redundancies such as those encountered with the short scan need to be carefully addressed to avoid artifacts. A common approach is the utilization of a Parker-like weight. In iterative reconstruction, there is a priori no need to use such a weight, but, if desired this weight can be introduced as part of the definition of C. In this section, we assume that all measurements have the same variance, and we study the impact of using a Parker-like weight for the definition of C versus using the identity matrix.

A. Data redundancy handling

To handle redundancies in the fan-beam data for the case of a short scan, we use a smooth weighting function

$$m(\lambda,\gamma) = \frac{c(\lambda)}{c(\lambda) + c(\lambda+\pi-2\gamma)}\ , \tag{5}$$

where

$$c(\lambda) = \begin{cases} \cos^2\left(\frac{\pi(\lambda-\lambda_s-d)}{2d}\right) & \text{if} \quad \lambda_s \le \lambda < \lambda_s + d \\ 1 & \text{if} \quad \lambda_s + d \le \lambda < \lambda_e - d \\ \cos^2\left(\frac{\pi(\lambda-\lambda_e+d)}{2d}\right) & \text{if} \quad \lambda_e - d \le \lambda < \lambda_e \end{cases} \tag{6}$$

with d being the angular interval over which $c(\lambda)$ smoothly drops from 1 to 0 [4]. If d is small, then the weighting function is similar to that of [5]. On the other hand, if d is large, then the weighting function is similar to Parker weighting [6].

B. Experiment details

In this second experiment, we investigate the influence of handling data redundancy in the Landweber algorithm. For this investigation, we created both full scan and short scan fan-beam data sets of the FORBILD head phantom. The data redundancy can be handled by setting the statistical matrix elements, C_{ij}^{-1}, equal to the result of Eqn. 5. We choose the angular interval $d = 0, 5, 30$ in the function $c(\lambda)$ in Eqn. 6, where $d = 0$ means using no weights since $m(\lambda,\gamma) = 1$. For that experiment, we created additionally 10 noisy realizations of each fan-beam data set.

1) Image quality: Image quality was assessed in terms of resolution, bias and noise properties. Our noise measurements include the square root of the pixel variance, $\overline{\sigma}$. All bias and noise metrics were computed for the reconstructions obtained every fifth iteration.

2) Resolution: The modulation transfer function (MTF) was used to evaluate resolution. This function was obtained using a phantom that consists only of the central low contrast ellipse within the FORBILD head phantom (see Fig. 6, area 2). For any reconstruction of this phantom, an edge profile that gives the reconstructed value as a function of the distance from the ellipse is computed. Then, the MTF is obtained as the Fourier transform of the differentiated edge profile. Due to our linear reconstruction method, this approach is suitable to evaluate the resolution achieved within area 2 in Fig. 6.

Since the resolution varies from one image representation to the other and also changes at a different pace for each representation, we present all our figures of merit as a function of the mean MTF value. To obtain the mean MTF value, we computed the area under the MTF curve over the range defined by the Nyquist frequency for the data. This was done for every fifth iteration up to 1000 iterations.

Fig. 5. The dark ring area was used for the calculation of the bias for the bowtie filter experiment.

Fig. 6. The hatched area 1 was used for the calculation of the bias since this area is not affected by edge artifacts. Area 2 was used for the MTF calculation.

3) Bias: Bias was evaluated as the mean reconstruction error over pixels located within the hatched white area 1 in Fig. 6 since this area is not affected by edge artifacts. The error for any given pixel was defined as the absolute difference between the reconstructed value and the true attenuation value of the hatched area, which is 50 HU.

4) Noise metrics: The image noise was analyzed over a region of interest (ROI) that corresponds to the hatched area 1 in Fig. 6. The noise magnitude, $\overline{\sigma}$, was evaluated from pixel variance computations. Using our 10 noise realizations, we first subtract the bias image from the noisy image. Then, the pixel variance was estimated for each pixel location in the ROI. Afterwards, the results were averaged over all pixels in the ROI. The square root of this mean was defined as $\overline{\sigma}$.

C. Results

Figure 7 shows the reconstructed images for the short scan case with an angular interval of $d = 0, 5, 30$ and for the full scan case without and with noise. The bias metric and the mean standard deviation as a function of the mean MTF value is shown in Fig. 8 and in Fig. 9.

Taking data redandancy into account helps to reduce image artifacts in reconstructed short scan images. However, this has some strong side effects (Tab. II). On the one hand, the bias in the image can be reduced by some HU so that less iterations are required to obtain good looking images but on the other hand, the noise level is negatively influenced by that.

scan modus	iteration number to reach the mean MTF value 0.70	absolute error [HU]	mean std value $\overline{\sigma}$
full	240	1.07	50.63
short, $d = 0$	265	6.21	52.95
short, $d = 5$	385	3.37	61.18
short, $d = 30$	360	1.98	60.89

TABLE II

V. Conclusions

We have shown in this work that utilizing the iteration number is rarely an effective means to regularize the reconstruction in x-ray CT imaging. As we have seen, both statistical weights and redundancies in the data set can easily introduce significant errors that differ from resolution errors and only dissapear after a large number of iterations. Whereas the resolution reached after say 250 iterations may be deemed satisfactory, the user will generally observe that the image quality is not. Hence, it is needed to iterate far beyond the desired resolution to first remove bias, and then post-smooth the result to attain the desired resolution. Under such circumstances, the penalized maximum-likelihood solution might be perceived as a more attractive reconstruction procedure. An alternative approach might be to initialize the reconstruction process with a filtered-backprojection procedure. However, in this case, it is important to understand what component of this first image remains when the reconstruction is completed.

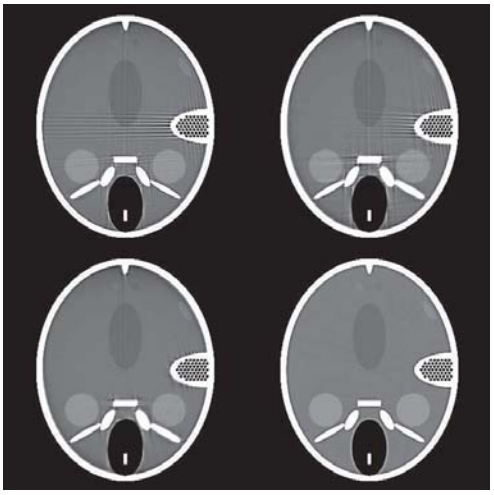


Fig. 7. Iterative reconstruction obtained from 250 iterations of the Landweber algorithm: (first row, left) in short scan geometry with $d = 0$, (first row, right) with $d = 5$, (second row, left) with $d = 30$, (second row, right) in full scan geometry. Grayscale: $[1, 1.1]$.

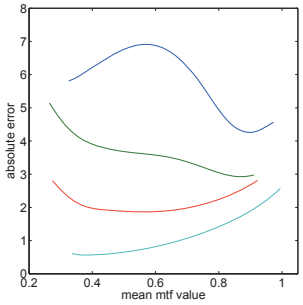


Fig. 8. Bias as a function of the mean MTF value obtained by varying the number of iterations by steps of 5 beginning with 50 iterations.

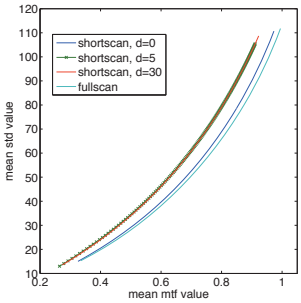


Fig. 9. Mean standard derivation as a function of the mean MTF value obtained by varying the number of iterations by steps of 5 beginning with 50 iterations.

References

[1] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in Medicine and Biology* **49**, pp. 2463 – 2474, 2004.

[2] G. H. Golub and C. F. van Loan, *Matrix Computations*, John Hopkins University Press, 1996.

[3] J. Hsieh, *Computed Tomography Principles, Design, Artifacts, and Recent Advances*, SPIE Optical Engineering Press, 2003.

[4] F. Noo, M. Defrise, R. Clackdoyle, and H. Kudo, "Image reconstruction from fan-beam projections on less than a short scan," *Phys. Med. Biol.* **47**, pp. 2525–46, 2002.

[5] G.-H. Chen, R. Tokalkanahalli, T. Zhuang, B. E. Nett, and J. Hsieh, "Development and evaluation of an exact fan-beam reconstruction algorithm using an equal weighting scheme via locally compensated filtered backprojection," *Med. Phys.* **33**, pp. 475–81, 2006.

[6] D. L. Parker, "Optimal short-scan convolution reconstruction for fan beam ct," *Med. Phys.* **9**, pp. 254–7, 1982.

Anatomically-guided MAP Reconstruction of Partial-ring TOF PET Data Using Spots-on-Smooth Image Representation Model

Tetsuya Kobayashi, Keishi Kitamura and Hiroyuki Kudo

***Abstract*—This paper investigates through simulation studies the image quality improvement in a partial-ring TOF PET scanner using an anatomically-guided MAP reconstruction method proposed by Kobayashi and Kudo [3, 4]. The reconstruction method adopts the special image representation model called the SOS (spots-on-smooth) model and estimates separately and simultaneously the two component images of the activity image, the piecewise-smooth background image representing a normal activity distribution and the sparse spot image representing an abnormal lesion image. The simulation results show that the MAP reconstruction method could provide better contrast-to-noise ratio than the conventional ML reconstruction.**

I. Introduction

ALONG with the popularization and technological maturation of the integrated PET-CT scanner, the center of research interest has moved to the development of integrated PET-MR scanners. In this situation, we are now developing a multi-modal compatible, stand-alone-type *flexible* PET scanner to expand the application area of PET. In our concepts, as shown in Fig. 1, the flexible PET scanner consists of *adjustable* two detector units and their support apparatus, and then scans a patient lying on a bed equipped by *other* medical instrumentation. The adjustability of the detector geometry is required to fit flexibly both various-sized and various-shaped patients and beds. For example, when there is an open space under the bed as shown in Fig. 1 and Fig. 2 (a), (b), it is possible to arrange the two detector units to sandwich a patient from both one's front-to-back and left-to-right directions; however, when there is no open space, the left-right-type or reverse U-shaped arrangement in Fig. 2(c) is forced. Here, the target instrumentation coupled with the flexible PET will not be restricted to the imaging devices such as CT, MR and SPECT, and includes also the non-imaging devices such as radiation therapy devices and surgical beds. Those can lead to the PET-image-guided radiation therapy and surgery. The advantage of the flexible PET would be not only in its multi-modal extensibility to the already-installed devices but also in its device cost due to the reduced number of detectors.

T. Kobayashi and K. Kitamura are with the Technology Research Laboratory, Shimadzu Corporation, Kyoto, Japan (e-mail: t_kobaya@shimadzu.co.jp).

H. Kudo is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan.

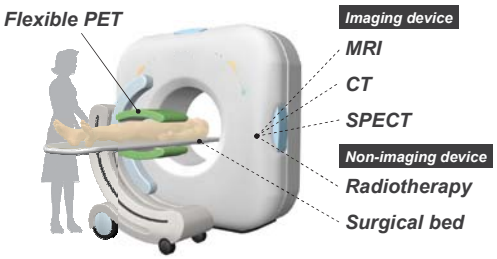


Fig. 1. A conceptual drawing of the flexible PET and candidate medical instrumentations coupled with the flexible PET.

Fig. 2. Various types of the partial-ring PET geometries.

However, as Fig. 1 and Fig .2 show, the flexible PET is a *partial-ring* scanner, i.e., not a conventional full-ring scanner, and its limited angular coverage causes missing line-of-responses (LORs) in the sinogram space. In such a partial-ring scanner having no time-of-flight (TOF) measurement capability, distortion- and artifact-free image reconstruction is not feasible without detector rotation for padding the sinogram space. However, Surti et al [1] showed that accurate TOF information improves reconstruction image quality in a dedicated breast partial-ring scanner without detector rotation. Recently, we also investigated through maximum-likelihood reconstruction studies the detector geometric and TOF requirements for the partial-ring whole-body scanner for obtaining comparable image quality to a full-ring non-TOF scanner [2]. The results showed that the partial-ring TOF scanner could provide comparable contrast-to-noise performance to the conventional non-TOF scanner under some measurement conditions. However, due to the missed spatial and statistical information in the partial-ring geometry, the geometry-dependent artifacts such as directional noise and edge blurring were observed.

In this study, to improve the image quality of a partial-ring TOF PET scanner, we applied the anatomically-guided MAP (maximum *a Posteriori*) reconstruction method [3, 4] dedicated to lesion detection and evaluated visually and quantitatively the lesion contrast and background noise performance.

II. RECONSTRUCTION METHOD

In this section, we explain briefly the reconstruction method on which we focus in this paper, the anatomically-guided MAP reconstruction method [3, 4], which is called the SOS (spots-on-smooth) -MAP.

A. Image Representation Model

In most of the iterative reconstruction methods for emission tomography, the activity distribution is represented by a single image vector **x** and, in case of the MAP reconstruction, the activity image **x** is regularized directly by some local smoothness penalty $U(\mathbf{x})$ on the assumption that the image **x** is locally smooth *everywhere*. However, this assumption is often violated by the presence of activity discontinuity then such penalty is unmatched to the real object. In contrast, in the SOS-MAP method, the activity image **x** is assumed to have *potentially* small, isolated lesions and represented by the sum of the *piecewise-smooth background image* **b** and the *sparse spot image* **s**, i.e., $\mathbf{x} = \mathbf{b} + \mathbf{s}$ (Fig. 3). In the other words, the images **b** and **s** represent the normal and abnormal components of the activity distribution, respectively. This image representation model called the spots-on-smooth (SOS) model allows ones to regularize *adaptively* the component images by different penalty functions.

Fig. 3. The SOS (spots-on-smooth) image representation model [3].

B. Problem Formulation

In the SOS-MAP method, the activity image **x** is reconstructed indirectly with $\mathbf{x}=\mathbf{b}+\mathbf{s}$ by solving the following optimization problem of estimating separately and simultaneously the background image **b** and the spot image **s**:

$$\begin{aligned} &\text{minimize } F(\mathbf{b},\mathbf{s}) \quad \text{subject to } \mathbf{b} \ge 0, \mathbf{b}+\mathbf{s} \ge 0 \\ &F(\mathbf{b},\mathbf{s}) = L(\mathbf{b}+\mathbf{s}) + \beta_b U_b(\mathbf{b}) + \beta_s U_s(\mathbf{s}) + \gamma D(\mathbf{s}) \end{aligned} \tag{1}$$

where $L(\mathbf{b}+\mathbf{s})$ is the well-known negative Poisson log-likelihood of the measured data, $U_b(\mathbf{b})$ and $U_s(\mathbf{s})$ are the local smoothness penalties for **b** and **s**, respectively, $D(\mathbf{s})$ is the sparseness penalty for **s**, and β_b, β_s and γ are the hyperparameters. The functional form of each regularization term is as follows [4]:

$$U_b(\mathbf{b}) = \sum\nolimits_{j\in\Omega} \left(b_j - m_j(\mathbf{b})\right)^2 / 2$$
$$\text{with } m_j(\mathbf{b}) = \sum\nolimits_{l\in N_j^b} \overline{w}_{jl} b_l, \quad \overline{w}_{jl} = w_{jl} \Big/ \sum\nolimits_{l'\in N_j^b} w_{jl'}, \tag{2}$$

$$U_s(\mathbf{s}) = \sum_{j\in\Omega}\sum_{l\in N_j^s} (s_j - s_l)^2 / (2d_{jl}), \tag{3}$$

$$D(\mathbf{s}) = \sum\nolimits_{j\in\Omega} |\, s_j \,| \tag{4}$$

where Ω is the set of all the pixel indexes, $m_j(\mathbf{b})$ is the j-th pixel value in the locally-smoothed image of **b**, N_j^b and N_j^s are the neighborhood regions centered on the j-th pixel in the background and spot images, respectively, w_{jl} is the smoothing weight assigned to the 2-pixel clique (j,l) in the background image, d_{jl} is the Euclidian distance between the j-th and l-th pixels. For preserving potential activity discontinuity in the background (normal) image, the value of the smoothing weight w_{jl} is defined with the assistance of the anatomical image (i.e., CT or MRI), **z**, as follows:

$$w_{jl} = h((z_j - z_l)/t) \Big/ \sum\nolimits_{l'\in N_j^b} h((z_j - z_{l'})/t) \tag{5}$$

where $h(\cdot)$ is Gaussian function with standard deviation of one and t is the parameter controlling the edge sensitivity for the anatomical image. The values of w_{jl} would be large if j-th and l-th pixels belong to the same anatomical region, and vice versa.

In the above definition, $U_b(\mathbf{b})$ is the weighted mean prior encouraging the *intra-region* local smoothness of **b**, $U_s(\mathbf{s})$ is the well-known quadratic prior encouraging the local smoothness of **s** and $D(\mathbf{s})$ is the L_1 norm prior (non-differentiable) encouraging the sparseness of **s**. The spatial size of N_j^b and the value of β_b are the crucial parameters for eliminating isolated spots from the background image, and the value of γ is the crucial *threshold-like* parameter for discriminating lesions (i.e., signal to preserve) and statistical fluctuation (i.e., noise to eliminate). The log-likelihood term and all the regularization terms in (1) are convex w.r.t. both **b** and **s**, and then the reconstruction problem (1) is a lower-bound-constrained, non-differentiable, convex optimization problem.

C. Iterative algorithm

We implemented the iterative algorithm [4] developed for solving the optimization problem (1). The algorithm updates the image variables **b** and **s** simultaneously (i.e. not alternately).

The derivation of the algorithm is based on the majorization-minimization technique [5] with De Pierro's additive and multiplicative convexity tricks [6,7] for constructing separable surrogate function for all the non-separable terms $L(\mathbf{b}+\mathbf{s})$, $U_b(\mathbf{b})$ and $U_s(\mathbf{s})$, and the soft-thresholding shrinkage operation [8] for handling the non-differentiability of the L_1 norm $D(\mathbf{s})$.

The computational complexity of the iterative algorithm is almost same level with the De Pierro's MAP-EM [7] because the algorithm mainly requires to compute (i) the gradient of $L(\mathbf{b}+\mathbf{s})$, $U_b(\mathbf{b})$ and $U_s(\mathbf{s})$ at each iteration and (ii) the curvature of the separable surrogate functions for $U_b(\mathbf{b})$ and $U_s(\mathbf{s})$, which are constant then pre-computable, and (iii) the next iterate using the explicitly-defined update formula (omitted for saving space), in which the soft-thresholding operation induced by the L_1-norm is implicitly embedded. This algorithm satisfies automatically the non-negativity constraint for $\mathbf{b} + \mathbf{s}$ due to the term $\log(b_j+s_j)$ appeared in the surrogate function. In [4], the initial estimates of **b** and **s** are set to the ML image and zero image, respectively. In this case, as iteration proceeds, the non-smooth component in **b** (i.e., isolated spots) is gradually disappeared then **b** gets smoother and, at the same time, the disappeared spots are recovered in **s** *if* its magnitude of intensity is large.

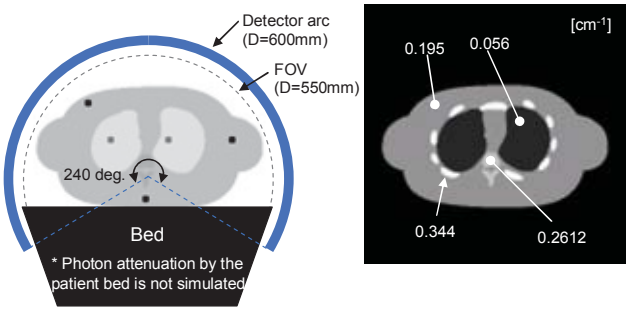
III. IMAGING SIMULATIONS

This section explains the simulation setup in the data

generation and image reconstruction experiments, and the method of image quality assessment.

A. Simulation Setup

In this study, we tested the reverse U-shaped partial-ring geometry in Fig. 4 (a), and compared with a full-ring geometry. The diameter of the detector arc/ring was 600 mm and the diameter of the reconstruction field-of-view (FOV) was 550 mm. In the partial-ring geometry, the central angle of *each* detector arc was 120 degree and then the total angular coverage was 240 degree. The timing resolution was 500ps (7.5 cm FWHM). Simulated data format was a 2-D TOF sinogram sized 128 [radial bin] × 128 [angular bin] × 10 [TOF bin]. The radial sampling interval was 4.5 mm. In the data generation process, the photon attenuation in the object was simulated but other physical factors (finite crystal size (detector response), scatter and random events, etc.) were not. The TOF sinogram was generated by the numerical convolution of the activity function along an LOR and the TOF kernel (the convolution of Gaussian function of 7.5 cm FWHM and a rectangular function corresponding to the spatial TOF bin).



(a) Simulated partial-ring geometry (b) Anatomical CT image used in the smoothing penalty $U_b(\mathbf{b})$

Fig. 4. (a) The simulated partial-ring geometry and (b) the anatomical CT image used in the SOS-MAP reconstruction.

We used as an activity distribution a human body phantom shown in Fig. 4 (a), sized 128 × 128 pixels and the pixel size is 4.5 mm. The phantom contains five hot lesions sized 13.0 mm in diameter. The lesion-to-background activity ratio was four for all the lesions. The total count in the *full-ring* geometry was set to 10^6 counts and the corresponding Poisson noise was added. Then, the total count in the partial-ring geometry having missing LORs was less than 10^6 counts, i.e., emission scan duration was assumed to be equal in both the full- and partial-ring geometries. Thirty noise realizations of sinogram were generated and reconstructed for the ensemble assessment of image quality.

Image reconstruction was performed with the AW (attenuation-weighted)-ML-EM [4] and the SOS-MAP for the partial-ring TOF PET data and with the AW-ML-EM for the full-ring non-TOF PET data. In the SOS-MAP, we tested several combinations of the hyperparameters (β_b, β_s, γ), and the other parameters (initial estimate, # of iteration, the size of N_j^b and N_j^s, the value of t) were set empirically. The parameter setting in the SOS-MAP was summarized in Table 1. Also, the noise-free CT image (70 keV) in Fig. 4 (b) was used as the anatomical image for computing the smoothing weights $\{w_{jl}\}$ in (5). We note that the anatomical image does not contain the lesions. The conventional anatomically-guided MAP methods [9, 10] often require lesion boundaries for achieving the lesion-preserving denoising but the SOS-MAP does not requires that because the lesion-preserving denoising is based on the sparseness, not smoothness.

Table 1. The parameter setting in the SOS-MAP

Initial estimate	Background : AW-ML-EM (100 iteration) Spot : zero image
# of iteration	300
β_b	$\{10^1, 10^2, 10^3, 10^4\}$
β_s	$\{1.0\times10^{-2}, 2.0\times10^{-2}, 4.0\times10^{-2}, 8.0\times10^{-2}\}$
γ	{0.1, 0.3, 0.5, 0.7}
N_j^b	Circle (radius: 3.5 pixels)
N_j^s	Rectangle (3×3 pixels)
t	0.03 cm^{-1}

B. Image Quality Assessment

Image quality was evaluated quantitatively with the contrast recovery coefficient (CRC) and the contrast-to-noise ratio (CNR) defied as follows:

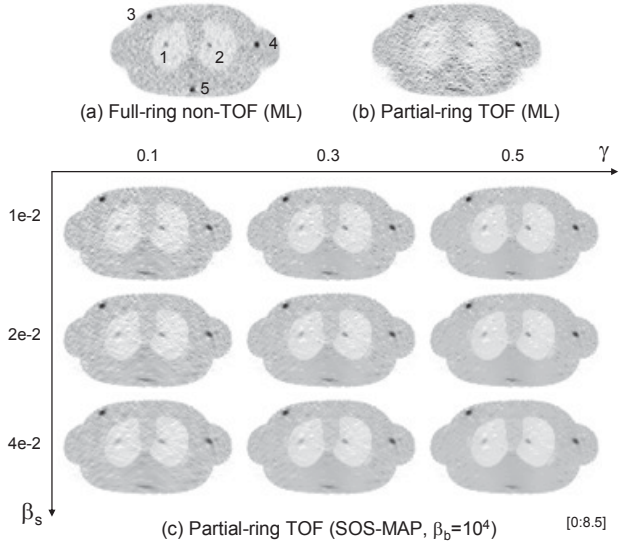
$$CRC(\%) = \left\langle \left(Mean_{SP} / Mean_{BG} - 1 \right) / \left(4 - 1 \right) \right\rangle \times 100 \qquad (6)$$

$$CNR = \left\langle \left| Mean_{SP} - Mean_{BG} \right| / \sqrt{\mathrm{Var}_{\mathrm{BG}}} \right\rangle \qquad (7)$$

where $Mean_{BG}$ and $Mean_{SP}$ are the mean reconstructed pixel values of the background and spot ROIs, respectively, and Var_{BG} is the variance of the reconstructed pixel values in the background ROI. The CRC and CNR are calculated for *each* lesion, and the background ROI is placed near each lesion.

IV. Results and Discussion

The sample reconstructed images by the AW-ML-EM (50 iter.) and the SOS-MAP are shown in Fig. 5. The images of the SOS-MAP method in Fig. 5 are the sum of the estimated background and spot images shown partially in Fig. 6.



(a) Full-ring non-TOF (ML) (b) Partial-ring TOF (ML)

(c) Partial-ring TOF (SOS-MAP, $\beta_b=10^4$)

Fig. 5. Sample reconstructed image of the different geometry, and reconstruction conditions. The image of the SOS-MAP method is the sum of the estimated background and spot images.

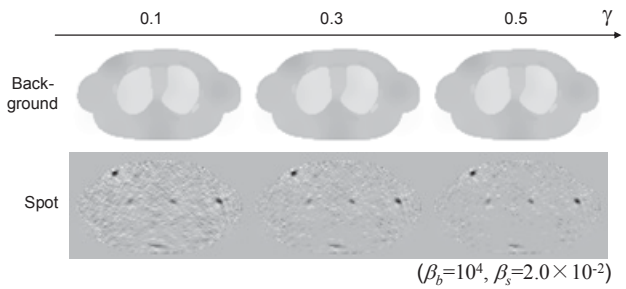
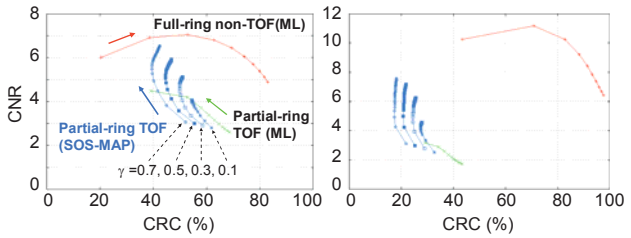


Fig. 6. Sample reconstructed background and spot images by the SOS-MAP method (β_b=10^4, β_s=2.0×10^{-2}).

The visual appearance of the sum image of the SOS-MAP was not so sensitive to the value of β_b in the range tested here so that only the results with β_b=10^4 are shown in Fig. 5 and Fig.6.

The CRC versus CNR plots for each geometry/reconstruction condition are shown in Fig. 7. Here, the plots for the leion-1 in the left lung and the lesion-5 near the back (i.e., the detector open space) are shown. The plot points are drawn at every ten iterations and the arrows in Fig. 7 (a) indicate the direction corresponding to the increase of the iteration number.



(a) Lesion-1 (left lung) (b) Lesion-5 (back)

Fig. 7. The CRC versus CNR plots for each geometry/reconstruction setting.

The ML reconstruction in the partial-ring TOF scanner in Fig. 5 (b) has a larger noise than that in the full-ring non-TOF scanner in Fig. 5 (b). The large noise is degrading the spot visibility, and also the directional noise and edge blurring are observed. This is apparently due to the missed spatial and statistical information in the partial-ring geometry. However, as Fig. 5 (c) shows, such image quality degradation is improved well by the SOS-MAP reconstruction with the proper selection of the hyperparameters. Particularly, in the SOS-MAP reconstruction, the image noise and the spot visibility were improved well by the combination of the smoothing effects of the terms $U(\mathbf{b})$ and $U(\mathbf{s})$ and the sparsifying effect of the term $D(\mathbf{s})$ and, at the same time, the edge blurring was reduced by the anatomically-guided (i.e., edge-preserving) smoothing effect of the term $U(\mathbf{b})$. Also, Fig. 6 shows that the separate reconstruction of the piecewise-smooth background image and the sparse spot image succeeded in the SOS-MAP method.

Fig. 7 shows that, in the partial-ring TOF reconstruction, the contrast recovery of the SOS-MAP method is lower than that of the ML reconstruction at the final iteration. This is unavoidable due to not only the smoothing effect of the term $U(\mathbf{s})$ but also the soft-thresholding *shrinkage* (i.e., intensity suppression) effect of the term $D(\mathbf{s})$. However, at the expense of the contrast recovery, the contrast-to-noise ratio, which is generally correlated with the lesion detectability, of the SOS-MAP method is better that that of the ML method due to its pretty low noise.

To improve the contrast recovery performance of the SOS-MAP reconstruction, the use of the L_0-norm prior as an alternative to the L_1-norm would be theoretically promising and effective, but practically the L_0-norm is non-convex, non-differentiable and non-continuous then the corresponding optimization problem becomes hard to solve. This is the reason why we choose the L_1-norm as a sparsifying penalty.

V. Conclusion

To improve the reconstruction image quality of the partial-ring TOF PET data, the performance of an anatomically-guided MAP reconstruction method called the SOS-MAP is evaluated. Thanks to the spots-on-smooth image representation model, the adaptive regularization (i.e., smoothing the background image and sparsifying the spot image) was introduced to the reconstruction. As a result, in the partial-ring geometry, the contrast-to-noise ratio was improved by the SOS-MAP method and it was better than that of the ML reconstruction. Also, due to the reduced statistical noise, directional artifact and edge blurring, the visual appearance was improved well.

References

[1] S. Surti and J.S. Karp, "Design considerations for a limited angle, dedicated breast, TOF PET scanner," Phys. Med. Biol., vol. 53, no. 11, pp. 2911-2921, 2008.

[2] T. Kobayashi and K. Kitamura, "Design Considerations for a Partial-Ring, Multi-Modal Compatible Whole-Body TOF PET Scanner: Flexible PET," *Conference record of 2012 IEEE NSS/MIC*, M11-6, 2012.

[3] T. Kobayashi and H. Kudo, "Fusion of image reconstruction and lesion detection using a Bayesian framework for PET/SPECT", *Conference record of 2008 IEEE NSS/MIC*, pp.3617-3624, 2008.

[4] T. Kobayashi, "PET image reconstruction using a priori knowledge" (in Japanese), Ph.D. dissertation, University of Tsukuba, 2011.

[5] K. Lange, D.R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. J. Comput. Graph. Stat., Vol. 9, pp. 1–20, 2000.

[6] A.R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. IEEE Trans. Med. Imaging, Vol. 12, pp. 328–333, 1993.

[7] A.R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Trans. Med. Imaging, Vol. 14, pp. 132–137, 1995.

[8] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm. Pure Appl. Math., Vol. 57, pp. 1413–1457, 2004.

[9] G. Gindi, M. Lee, A. Rangarajan, and I.G. Zubal. Bayesian reconstruction of functional images using registered anatomical images as priors. Lect. Notes Comput. Sci., Vol. 511, pp. 121–131, 1991.

[10] C. Comtat, P.E. Kinahan, J.A. Fessler, T. Beyer, D.W. Townsend, M. Defrise, and C. Michel. Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels. Phys. Med. Biol., Vol. 47, pp. 1–20, 2002.

Statistical Interior Tomography via Dictionary Learning without Assuming an Object Support

Junfeng Wu, Xuanqin Mou, Hengyong Yu and Ge Wang

Abstract—**In this article, we propose a new dictionary learning approach for CT reconstruction. Specifically, dictionary learning is incorporated into the statistical iterative reconstruction (SIR) framework for interior tomography. As a novel prior, the zeroth order image moment is utilized in the reconstruction framework. The objective function is minimized using an alternating scheme. Preliminary experimental results demonstrate that the proposed approach produce excellent interior tomography results.**

Index Terms—**Dictionary learning, Direct current component, Interior tomography, Statistical iterative reconstruction**

I. Introduction

IT is well-known that interior problem does not have a unique solution in an unconstrained image space even if we know an object support (OS) exactly [1]. Recently, it has been proven that if a sub-region in an internal region of interest (ROI) is known the ROI can be reconstructed using the differentiated backprojection (DBP) projection on convex set (POCS) technique [2-4] or differentiated backprojection (DBP) singular value decomposition (SVD) method [5, 6]. It has been also [5] proposed that if the ROI is piecewise constant [7, 8] or polynomial [9, 10], the compressive sensing based approach such as the total variation (TV) minimization or high order TV (HOT) minimization can solve the interior problem exactly and stably. Then, statistical iterative reconstruction (SIR) with total variation (TV) regularization [11] has been proposed to solve the interior problem with performance superior to the previous counterparts, especially in the cases of low-dose and few-view settings.

A very recent study [12] shows that introducing dictionary learning (DL) and sparse representation technique into SIR framework is promising for low-dose CT reconstruction. In particular, the DL-based regularization term is superior to the TV-based candidate in terms of preserving image details and removing artifacts. Encouraged by this finding, here we investigate the DL-based SIR framework for interior tomography.

One crucial issue in interior tomography is intensities drop around the peripheral region of an ROI. When priori information does not exist or is not sufficiently strong, image intensity could deviate dramatically from the truth near the boundary of the ROI. There are a number of ways to suppress this problem by introducing additional information about the object to be reconstructed. Currently, an effective mean is to enforce an object support (OS). However, this is not always available in practical scenarios. Hence, we hypothesize that the zeroth order image moment is a surrogate, which can be easily measured in the projection domain [16].

The main idea is to use the zeroth order image moment as a constraint in the DL-based SIR framework for faithful interior reconstruction without an assumption on an OS. The rest of this paper is organized as follows. Section 2 introduces the proposed method. Section 3 presents the experimental results. Finally, Section 4 discusses the results and concludes the paper.

II. Methods

A. Statistical Iterative Reconstruction

SIR [13-15] is used for a long time in the CT field, due to the solid base of physical modeling and the ability to incorporate priori information. In SIR, an image is reconstructed by minimizing the following cost function:

$$\min_{\boldsymbol{\mu}} \left\{ \sum_{i=1}^{I} \frac{w_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - p_i)^2 + \beta U(\boldsymbol{\mu}) \right\}, \quad (1)$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_J)^T$ represents an image vector representing the attenuation coefficients of an object, $\mathbf{P} = (p_1, \ldots, p_I)^T$ denotes CT projection data, w_i is for statistical weighting associated with the ith ray, and $\mathbf{A}$ is the $I \times J$ system matrix. The forward projection operation is $[\mathbf{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{J} a_{ij}\mu_j$. $U(\boldsymbol{\mu})$ is a regularization term expressing the prior knowledge of the object, and β is a regularization parameter. The symbol T denotes the transpose operator. SIR with TV regularization has been proposed to solve the interior problem in the low-count situation [11].

This work is partially supported by NSFC through Grant No.61172163 and the Research Fund for the Doctoral Program of Higher Education of China through Grant No. 20110201110011.

J .F. Wu and X. Q. Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an, Shannxi 710049, China. (E-mail: wujf2004@163com, xqmou@ mail.xjtu.edu.cn).

H. Y. Yu is with Radiologic Science Division, Wake Forest University Health Sciences, Winston-Salem, NC, 27157, USA. (E-mail: hengyong-yu@ieee.org).

G. Wang is with Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA (E-mail: ge-wang@ieee.org).

B. Sparse Representation based on Learned Dictionary

Let a binary matrix $\mathbf{R}_s = \{r_{nj}^s\} \in \mathrm{R}^{N\times J}$ be an operator that extracts the s-th patch of $\boldsymbol{\mu}$, and $\mathbf{R}_s\boldsymbol{\mu}$ is the vector representation of a square 2D image patch of $\sqrt{N}\times\sqrt{N}$ pixels. $\mathbf{D}\in R^{N\times M}$ is a dictionary of M atoms represented in columns $d_k \in R^N$. Then, the sparse representation of $\mathbf{R}_s\boldsymbol{\mu}$ with $\mathbf{D}$ is to find a sparse vector $\boldsymbol{\alpha} = [\alpha_1;\ldots;\alpha_M]$ (i.e., most of the coefficients in α are zero or insignificant) such that $\mathbf{R}_s\boldsymbol{\mu} \approx \mathbf{D}\boldsymbol{\alpha}$. The sparse vector $\boldsymbol{\alpha}$ can be obtained by solving the following minimization problem:

$$\min_{\boldsymbol{\alpha}} \left\{ \left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha} \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_1 \right\}, \tag{2}$$

where λ is a regularization parameter.

An ROI image can be reconstructed by minimizing the following optimization problem:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\mu}} \left\{ \sum_{i=1}^{I} \frac{w_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - p_i)^2 + \beta \sum_{s=1}^{S} (\left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_s \right\|_1) \right\}. \tag{3}$$

C. Direct Current Priors

According to our previous work [16], the zeroth order Helgason–Ludwig consistency condition (HLCC) applied for equi-angular fan beam geometry can be expressed in the following equation:

$$\int_{-\pi/2}^{+\pi/2} p(\xi, \psi + \xi - \pi/2) E \cos\xi d\xi = m_{0,0} \tag{4}$$

where E is the distance from an x-ray to the rotation center, and $m_{0,0}$ is the zeroth order image moment, namely the sum of the image pixel values (which is the so-called direct current (DC) component). The variables ξ and ψ are the angle of an x-ray relative to the central x-ray and a projection view angle respectively. After computing the DC value, denoted as C in the following, of the image to be reconstructed using Eq. (4), we can incorporate the difference between the measured/assumed and reconstructed results into the objective function Eq. (3). In this study, we consider C as a prior estimated from clinical CT images, which can be also directly measured with one or more complete fan-beam views. Our experiments show that the accuracy of the estimated C is not critical, which will be detailed in Sections III and IV.

D. Proposed algorithm

The SIRDL with the DC prior can be formulated as

$$\min_{\boldsymbol{\mu},\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{I} \frac{w_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - p_i)^2 + \beta \sum_{s=1}^{S} (\left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_s \right\|_1) + \gamma \left\| \sum_{j=1}^{J} \mu_j - C \right\|_2^2 \right\} \tag{5}$$

The optimization of the above objective function can be implemented in the following three steps in an alternating fashion to meet some stopping criteria.

1) Sparse representation

$$\min_{\boldsymbol{\alpha}} \left\{ \sum_{s=1}^{S} (\left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_s \right\|_1) \right\}, \tag{6}$$

Eq. (6) is equivalent to the following convex constrained optimization problem:

$$\min_{\boldsymbol{\alpha}} \left\| \boldsymbol{\alpha}_s \right\|_1 \quad s.t. \quad \left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s \right\|_2^2 \le \varepsilon, \tag{7}$$

where ε is a nonnegative real parameter. In this project, the LASSO algorithm [17] is applied to solve Eq. (7).

2) Image updating

$$\min_{\boldsymbol{\mu}} \left\{ \sum_{i=1}^{I} \frac{w_i}{2} ([\mathbf{A}\boldsymbol{\mu}]_i - p_i)^2 + \beta \sum_{s=1}^{S} \left\| \mathbf{R}_s\boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}_s \right\|_2^2 \right\}, \tag{8}$$

Using the separable paraboliod surrogate method [13], Eq. (8) can be solved in the following iteration

$$\tilde{\mu}_j^k = \mu_j^{k-1} - \frac{\sum_{i=1}^{I} \left(a_{ij} w_i \left([\mathbf{A}\boldsymbol{\mu}]_i - p_i \right) \right) + 2\beta \sum_{s} \sum_{n=1}^{N} r_{nj}^s \left([\mathbf{R}_s\boldsymbol{\mu}]_n - [\mathbf{D}\boldsymbol{\alpha}_s]_n \right)}{\sum_{i=1}^{I} \left(a_{ij} w_i \sum_{m=1}^{J} a_{im} \right) + 2\beta \sum_{s} \sum_{n=1}^{N} r_{nj}^s \sum_{m=1}^{J} r_{nm}^s}, \tag{9}$$

$j = 1,\ldots J$, where the superscript $k = 1,\ldots K$ represents the iteration number.

3) Updating the reconstructed image with the DC prior

$$\mu_j^k = \tilde{\mu}_j^k - 2\gamma \left(\sum_{j=1}^{J} \tilde{\mu}_j^k - C \right) \tag{10}$$

The workflow of the proposed algorithm (SIRDL+HL) is summarized in Table 1

TABLE I. THE WORKFLOW OF SIRDL+DC

Given $\varepsilon > 0$ β, γ, C and K;

Initialize $\boldsymbol{\mu}^0$ and set k=0;

Step 1. Solve the optimization problem Eq. (7) for $\boldsymbol{\alpha}^k$ using the LASSO algorithm;

Step 2. Update the reconstructed image $\tilde{\boldsymbol{\mu}}^k$ using Eq. (9);

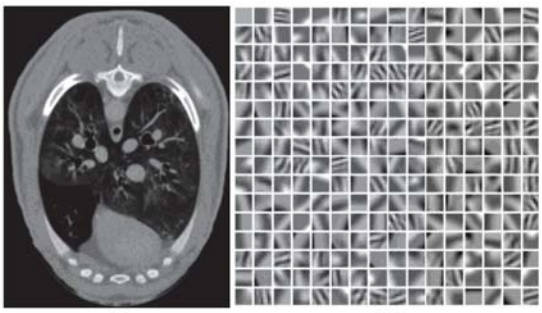
Step 3. Update the reconstructed image $\boldsymbol{\mu}^k$ using Eq. (10);

Until the stop criteria are met.

III. RESULTS

A sheep chest was scanned with both normal (100kV, 150mAs) and low does (80kV, 17mAs) protocols on a SIEMENS Somatom Sensation 64-Slice CT Scanner using a circular cone-beam scanning mode. The sinograms of the central slices are extracted, which are in fan-beam geometry. The radius of the trajectory is 57cm. 1160 projections are uniformly collected over a 360° range. For each projection, 672 detector elements are equi-angularly distributed to define a field of view (FOV) of 50.1 cm in diameter.

A baseline image in a 768×768 matrix covering a 43.63×43.63 cm^2 region is first reconstructed from a normal-dose sinogram using the classic FBP algorithm. Then the entire lung region of 495×370 pixels is extracted from the baseline image. Finally, a global dictionary $\mathbf{D} \in R^{64*256}$ with 256 atoms is trained



(a) (b)

Fig.1. Reconstructed baseline image from a normal-dose sinogram with a display window [-700, 800] HU (a) and the trained global dictionary consisting of 256 atoms (b).

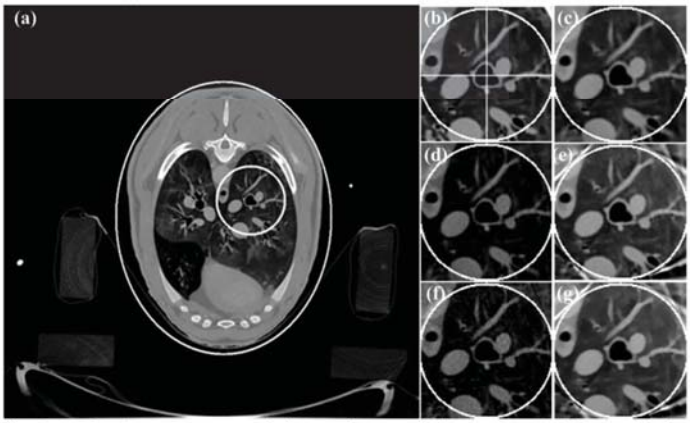
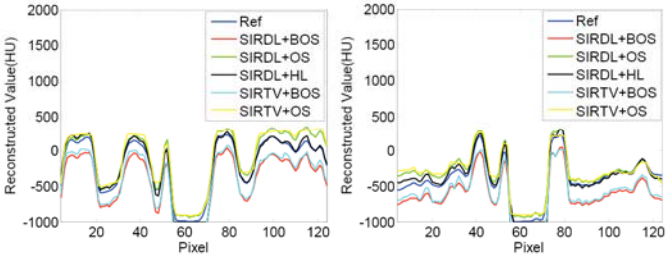


Fig.2. (a) is the reconstructed image from a full-scan normal-dose global dataset using the SIRDL and (b) is the magnified ROI. (c)-(g) are the reconstructed ROI images from truncated local projections using the proposed algorithm (SIRDL+HL), the SIRDL without OS knowledge(SIRDL+BOS), the SIRDL with exact OS (SIRDL+OS) , the SIRTV without OS knowledge (SIRTV+BOS) and SIRTV with exact OS (SIRDL+OS), respectively. The display window is [-700,800] HU.



(a) (b)

Fig.3. Profiles along the (a) horizontal and (b) vertical central lines of the ROIs in Fig. 2. The lines are indicated by the white crossing in Fig.2 (b).

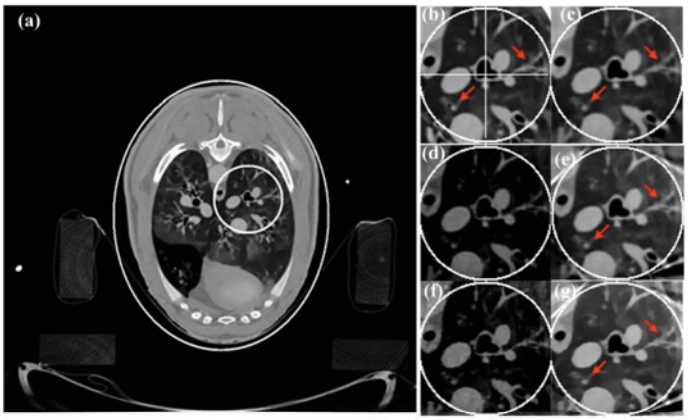
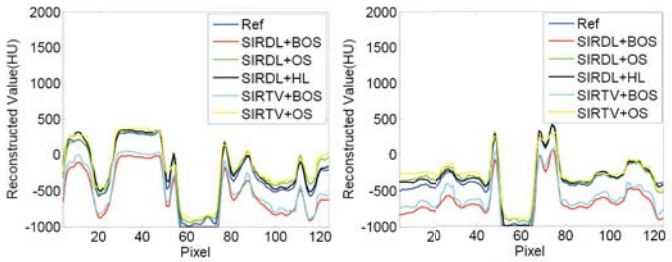


Fig.4. Same as fig. 2 but reconstructed from a low-dose sinogram.



(a) (b)

Fig.5 Profiles along the (a) horizontal and (b) vertical central lines of the ROI images in Fig.4.

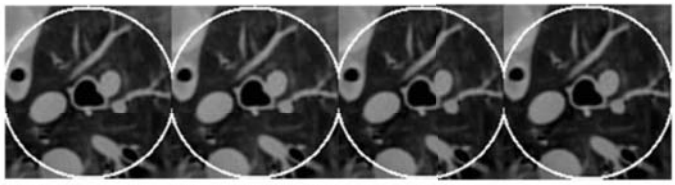
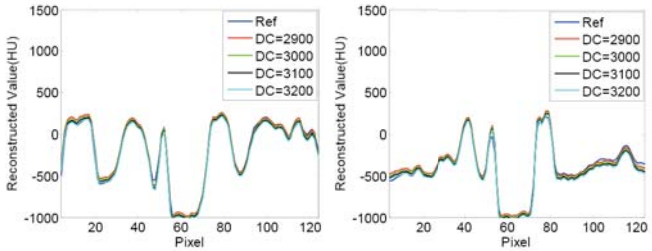


Fig.6. The reconstructed ROI images using the proposed algorithm from normal-dose dataset with different value of DC. From left to right, the DC values are 2900, 3000, 3100 and 3200, respectively. The display window is [-700,800]HU.



(a) (b)

Fig.7. Profiles along the (a) horizontal and (b) vertical central lines of the ROI images in Fig.4.

TABLE I

RMSE COMPARISONS OF THE RECONSTRUCTED ROI IMAGES FROM TRUNCATED NORMAL-DOSE DATA USING DIFFERENT ALGORITHMS (HU)

	SIRDL+BOS	SIRDL+OS	SIRDL+HL	SIRTV+BOS	SIRTV+OS
RMSE	2.2135e+003	831.8282	140.3900	1.9150e+003	1.0006e+003

TABLE II

RMSE COMPARISONS OF THE RECONSTRUCTED ROI IMAGES FROM TRUNCATED LOW- DOSE DATA USING DIFFERENT ALGORITHMS (HU)

	SIRDL+BOS	SIRDL+OS	SIRDL+HL	SIRTV+BOS	SIRTV+OS
RMSE	3.0150e+003	601.6093	346.9616	2.4030e+003	950.7038

using the online dictionary learning technique from the baseline image of the lung region [18]. The normal-dose FBP image and the trained dictionary are shown in Fig.1.To get a better sparse representation, the baseline image and the image to be reconstructed is preferred to be the same body part in most situations.

The whole image is firstly reconstructed using the SIRDL method from the normal-dose full-scan dataset, which will be used as reference to evaluate the proposed algorithm. Using the fan-beam HLCC, we obtain the value of DC as 2.9965e+003. To mimic an interior scan, the data are truncated manually to keep those passing through a selected ROI. Using the learned dictionary and the DC value, we reconstruct the ROI by the proposed algorithm from the normal-dose truncated local

projection data. The parameters are set to $\varepsilon = 6.25\text{e}-5$, $\beta = 0.015$ and $\gamma = 1.2716\text{e-}007$. The iterative process is stopped after 50 iterations with 40 subsets. Fig.2 (c) shows the reconstructed result.

For comparison analysis, the ROI is also reconstructed using the SIRDL without a known OS (SIRDL+BOS), SIRDL with a tight OS (SIRDL+OS), SIRTV without a known OS (SIRTV+BOS) and SIRTV with a tight OS (SIRTV+OS). The details of the SIRTV algorithm can be found in [11]. The results are shown in Figs. 2. The representative profiles along the horizontal and vertical lines are shown in Fig. 3. The above experiments are repeated with a low-dose sinogram using the same parameters except $\beta = 0.02$, and corresponding results are shown in Figs. 4 and 5. Although both the SIRDL and SIRTV based methods outperform the classical FBP algorithm, the SIRDL based algorithms are superior to the SIRTV based ones in terms of recovering detail structures, especially for the low dose case. As shown in Fig. 4, there are some blocky structures and some detail structures are missed which are indicated by red arrows. Regarding to the issue of intensity drop, it can be observed that the reconstructed ROIs have severe DC-shift artifact without the knowledge of OS for the cases of both SIRTV and SIRDL, and the reconstructed ROIs have also a little DC-shift artifact even we know the exact OS. Contrastively, the SIRDL+HL recover the DC component very well.

To quantitatively analyze the reconstruction accuracy of recovering the DC component, we apply a Gaussian filter (G) with a scale factor of 2 to the reconstructed images and then compute the root mean square error as:

$$\text{RMSE} = \sqrt{(\boldsymbol{\mu}^r \otimes G - \boldsymbol{\mu}^* \otimes G)^2 / M} \qquad (11)$$

where $\boldsymbol{\mu}^r$ is the test image, $\boldsymbol{\mu}^*$ is the reference image and M is the total number of pixels inside the ROI. The calculated RMSEs for the images in Figs. 2 and 4 are listed in Tables I and II, respectively. It can be also observed that the proposed algorithm outperforms other algorithms in terms of reducing the DC-shift artifact.

To investigate the effect of the accuracy of estimated DC value on the recovery of intensity drop, the reconstructed ROIs form normal-dose sinogram with four different DC values around the true value are shown in Fig.6 and the corresponding central horizontal and vertical profiles are shown in Fig.7. It can be seen that the proposed algorithm is robust with respect to the DC value.

IV. Discussion and Conclusion

This work has shown that the SIRDL-based reconstruction strategy is superior to the SIRTV based counterpart for interior tomography, especially in the case of low-dose data. Even with the exact knowledge of OS, the ROI reconstruction using a CS based algorithm may produce a substantial DC-shift. Incorporating the constraint in terms of the DC value into the objective function of the SIRDL framework can well address this problem. The results have also indicated that the accuracy of the DC value does not affect the interior reconstruction much. As an interesting future topic, we will estimate the DC value of an image to be reconstructed from truncated projections.

In conclusion, we have incorporated the sparse representation in terms of learned dictionary and the constraint in terms of an image DC value into the SIR framework for interior tomography. Experimental results have demonstrated that our proposed approach reduces the DC-artifact effectively, and the DL-based algorithm outperforms the TV-based method in preserving fine structures, especially in a low-dose situation.

References

[1] F. Natterer, The mathematics of computerized tomography vol. 32: Society for Industrial Mathematics, 2001.

[2] M. Courdurier, F. Noo, M. Defrise, and H. Kudo, "Solving the interior problem of computed tomography using a priori knowledge," Inverse Problems, vol. 24, p. 065001, 2008.

[3] H. Kudo, M. Courdurier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," Physics in Medicine and Biology, vol. 53, p. 2207, 2008.

[4] Y. Ye, H. Yu, and G. Wang, "Exact interior reconstruction from truncated limited-angle projection data," Journal of Biomedical Imaging, vol. 2008, p. 5, 2008.

[5] H. Yu, Y. Ye, and G. Wang, "Interior reconstruction using the truncated Hilbert transform via singular value decomposition," Journal of X-Ray Science and Technology, vol. 16, pp. 243-251, 2008.

[6] A. Katsevich, "Singular value decomposition for the truncated Hilbert transform," Inverse Problems, vol. 26, p. 115011, 2010.

[7] H. Yu and G. Wang, "Compressed sensing based interior tomography," Physics in Medicine and Biology, vol. 54, p. 2791, 2009.

[8] H. Yu, J. Yang, M. Jiang, and G. Wang, "Supplemental analysis on compressed sensing based interior tomography," Physics in Medicine and Biology, vol. 54, p. N425, 2009.

[9] J. Yang, H. Yu, M. Jiang, and G. Wang, "High-order total variation minimization for interior tomography," Inverse Problems, vol. 26, p. 035013, 2010.

[10] E. Katsevich, A. Katsevich, and G. Wang, "Stability of the interior problem with polynomial attenuation in the region of interest," Inverse Problems, vol. 28, p. 065022, 2012.

[11] Q. Xu, X. Mou, G. Wang, J. Sieren, E. A. Hoffman, and H. Yu, "Statistical interior tomography," Medical Imaging, IEEE Transactions on, vol. 30, pp. 1116-1128, 2011.

[12] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-Dose X-ray CT Reconstruction via Dictionary Learning," Medical Imaging, IEEE Transactions on, vol. 31, pp. 1682-1697, 2012.

[13] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," Medical Imaging, IEEE Transactions on, vol. 21, pp. 89-99, 2002.

[14] J. Tang, B. E. Nett, and G. H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," Physics in Medicine and Biology, vol. 54, p. 5781, 2009.

[15] E. A. Rashed and H. Kudo, "Statistical image reconstruction from limited projection data with intensity priors," Physics in Medicine and Biology, vol. 57, p. 2039, 2012.

[16] H. Y. Yu, X. Q. Mou and Y. L. Cai, "Calibration of cone beam rotational X-ray image sequence," Chin. J. Electron., vol. 13, p. 5, 2004.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," The Journal of Machine Learning Research, vol. 11, pp. 19-60, 2010.

[18] http://spams-devel.gforge.inria.fr/.

Investigation of Template Structure for a Cone-Beam CT Signal Detection Task

Adrian A Sanchez[1], Emil Y Sidky[1], and Xiaochuan Pan[1,2]
[1]The University of Chicago, Department of Radiology, Chicago, IL 60637
[2]The University of Chicago, Department of Radiation and Cellular Oncology, Chicago IL 60637

***Abstract*—Task-based metrics of image quality provide valuable insight into the utility of images in CT. However the computation of these metrics can be intensive and rife with complications which arise from the properties of the reconstruction operation. Here, we investigate the computation of regularized forms of the 3D Hotelling template and the impact of regularization on SNR of the model observer as well as on computational efficiency. Results are presented for detection of an ellipsoidal signal with size on the order of a detector element, meant to mimic microcalcifications in cone-beam breast CT. We hypothesize that, through Tikhonov regularization in the solution of the 3D Hotelling template, structurally simple templates can be constructed whose performance mimics that of human observers, while simultaneously easing the computational burden inherent in solving for 3D Hotelling observer performance.**

I. Introduction

The plethora of image quality metrics which exists can lead to a difficult choice when seeking metrics to aid in imaging system design, including the image reconstruction algorithm. Task-based evaluation arguably provides the most meaningful metrics [1], [2], but application of these metrics to tomographic imaging comes with certain difficulties and necessary considerations not encountered in other forms of medical imaging [3]. For instance, the noise in a reconstructed image can have a high degree of correlation and is often sensitive to small implementation or design differences in the reconstruction algorithm. Therefore, in order to make full use of task-based metrics, a detailed investigation of the impact of the reconstruction algorithm on the formulation of these metrics is warranted. Researchers such as those in [4] have performed some investigations of channelized Hotelling observers (CHOs), but in the current work, we shall focus on computing ideal observer performance, rather than attempt to estimate human observer performance. This provides a clear upper bound on the performance of any human or mathematical observer.

In this work, we investigate the case of circular cone-beam CT reconstruction when the reconstruction operation can be represented as a linear discrete-to-discrete (matrix) operator. We compute the 3D Hotelling observer SNR (SNR_{Hot}) for a signal detection task. Further, we discuss the non-uniqueness of the Hotelling template in this case and demonstrate the use of Tikhonov regularization in solving for a unique template which approximates the behavior of the 3D Hotelling template while obtaining a structure which more closely resembles the signal of interest. We hypothesize that templates which most closely resemble the reconstructed signal could be predictive of human observer performance.

II. Background

A. Linear Reconstruction Algorithm

The particular image reconstruction algorithm used in this study is the back-projection filtration (BPF) image reconstruction algorithm developed by our group [5]. In this algorithm, the imaging volume is decomposed into individual chords of the scanning trajectory, along which the reconstruction takes place. Although this algorithm is developed based on a continuous-to-continuous model, in practice the reconstruction is performed using discretized versions of the various continuous inversion operations that comprise the algorithm. In general terms, we consider a linear image reconstruction algorithm A that takes a discrete set of data $\mathbf{g}$ and produces a discrete representation of an image in the form of voxel coefficients $\mathbf{y}$:

$$\mathbf{y} = A\mathbf{g} \tag{1}$$

where bold denotes quantities which are statistically variable. As the BPF algorithm consists of many linear processing steps, it is useful to consider A as the product of matrices representing each processing step:

$$A = \prod_i A_i. \tag{2}$$

The BPF algorithm can be said to consist of eight distinct processing steps as follows:
(1) Derivative filtration of the projection data
(2) Back-projection onto the chords comprising the image volume
(3) Computation of boundary terms for back-projection
(4) Weighting of chord profiles
(5) Inverse Hilbert transform of chord profiles
(6) Evaluation and addition of constant offset to chord profiles
(7) Inverse weighting of chord profiles
(8) Interpolation onto 3D Cartesian grid.
In our case, the final step is trivial, as we consider reconstruction onto a set of chords defined on a Cartesian grid, but it is included here for completeness.

It should be emphasized that A can be treated as a matrix. Therefore, if one considers the data covariance matrix K_g whose (i,j)th entry is defined by

$$(K_g)_{i,j} = Cov\left(\mathbf{g}_i, \mathbf{g}_j\right) \tag{3}$$

this representation of the reconstruction algorithm can be used to construct the image voxel covariance matrix K_y as

$$K_y = AK_gA^T \tag{4}$$

where the superscript T denotes the matrix transpose.

B. Binary Detection

The task we consider is a signal detection task. This task is a binary decision in which an observer is asked to view an image and to make a decision between two hypotheses: signal absent or signal present. In order to represent these hypotheses mathematically, consider a continuous object function $f(\vec{r})$ defined for $\vec{r} \in \mathbb{R}^3$. We then denote the continuous-to-discrete forward projection operation as the action of the operator $\mathcal{P}$, which maps the continuous object function $f(\vec{r})$ to the discrete data vector $\mathbf{g}$:

$$\mathbf{g} = \mathcal{P}f + \mathbf{n} \tag{5}$$

where $\mathbf{n}$ is an additive measurement noise term in the projection data domain. We consider the different elements $\mathbf{n}_i$ of the vector $\mathbf{g}$ to be independent and drawn from identical Gaussian distributions.

We shall label the signal-absent and signal-present hypotheses as H_0 and H_1, respectively. Since we consider detection in the reconstructed image domain, we must apply the reconstruction operator A to the data vectors representing each hypothesis. The resulting hypotheses may then be expressed mathematically as

$$\begin{aligned} H_0 \quad &: \quad \mathbf{y} = A\left(\mathcal{P}f_b + \mathbf{n}\right) \\ H_1 \quad &: \quad \mathbf{y} = A\left(\mathcal{P}\left(f_b + f_s\right) + \mathbf{n}\right) \end{aligned} \tag{6}$$

where f_b and f_s represent the background object and signal object, respectively. The task we consider is a signal-known-exactly, background-known-exactly (SKE/BKE) task, meaning that the observer has full knowledge of the reconstructed signal and the image noise statistics.

C. 3D Hotelling Observer

The Hotelling observer (HO) is the optimal linear mathematical model observer; i.e. the HO uses an optimal method of linearly combining the elements of $\mathbf{y}$ to make a decision between signal-absent and signal-present [1]. The Hotelling SNR is the figure of merit used in this work, and its computation is performed as follows: First, one must solve for w, the 3D Hotelling template, which is the image with which the HO masks a reconstructed image $\mathbf{y}$ in order to make the decision between signal-absent and signal-present. w is found by solving the linear equation

$$K_y w = \Delta\bar{y}, \tag{7}$$

where K_y is the image covariance matrix and $\Delta\bar{y}$ is the mean difference between the hypotheses H_0 and H_1 of Eqn. 6. Since the only source of statistical variability in Eqn. 6 is the additive noise term, the quantity $\Delta\bar{y}$ can simply be computed as

$$\Delta\bar{y} = \left(A\mathcal{P}\left(f_b + f_s\right) + A\bar{n}\right) - \left(A\mathcal{P}f_b + A\bar{n}\right) = A\mathcal{P}f_s, \tag{8}$$

where $\bar{n}$ is a vector whose elements represent the mean detector element noise. Note that the quantities f_b and f_s are not statistically variable in this case, as the task chosen is a signal-known-exactly / background-known-exactly (SKE/BKE) task. Eqn. 8, therefore, contains no statistically variable quantities.

Although the form of Eqn. 7 is compact, obtaining a direct solution for w is computationally nontrivial, since the matrix K_y can be too large to be stored in computer memory and is non-diagonal. In fact, the computational burden of solving Eqn. 7 is the basis for much active research. Here, we restrict ourselves to a small system so that direct matrix inversion is still feasible and the matrix K_y can be stored directly in computer memory, however in general iterative methods must be applied to Eqn. 7 in order to obtain the 3D Hotelling template.

Finally, after obtaining a solution to Eqn. 7, the HO SNR for a signal-detection task is computed by applying the Hotelling template w to the mean difference between the image vectors for the hypotheses H_0 and H_1, $\Delta\bar{y}$:

$$\text{SNR}^2_{Hot} = w^T\Delta\bar{y}. \tag{9}$$

where $\Delta\bar{y}$ is again given by Eqn. 8. The calculated HO performance can then provide an absolute upper bound on measured human observer performance.

One complication which arises in the case of applying the metric given in Eqn. 9 to computed tomography is that the matrix A defined by the reconstruction algorithm has a non-trivial nullspace. In order to illustrate this nullspace, the singular value decomposition of the reconstruction algorithm used in this work was computed, and the singular value spectrum is shown in Figure 1. Double precision floats were used, so that any singular value on the order of 10^{-15} or less is essentially equal to zero. The reconstruction parameters used to construct A in this case are discussed in section III-A1.

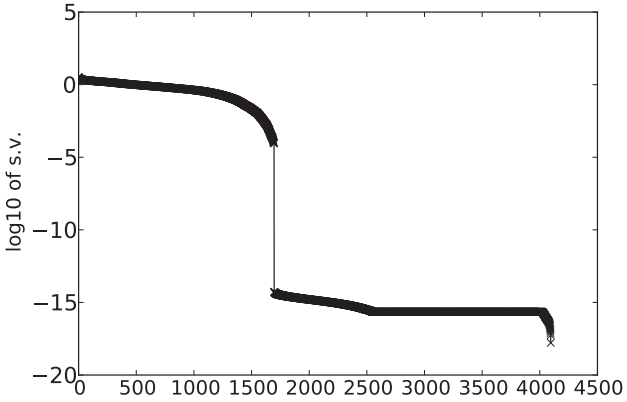


Fig. 1. The singular value spectrum of the reconstruction operator A corresponding to the BPF algorithm. Note that of the roughly 4100 singular values, approximately 1600 are zero, to within numerical precision.

One crucial consequence of the nullspace of A is that the matrix K_y becomes singular. This implies that the 3D Hotelling template w is no longer unique.

In order to enable us to investigate the effects of template structure on non-ideal model observer detectability, we have investigated the application of Tikhonov regularization to the solution of Eqn. 7. Through iterative methods, we have solved the related minimization problem defined by

$$w = \arg\min_x \left\{ ||K_y x - \Delta\bar{y}||^2 + \lambda^2 ||x||^2 \right\} \qquad (10)$$

for various values of the damping parameter λ. The problem defined in (10) is guaranteed to have a single, unique solution. In addition to the appeal of a unique solution to (10), there is a further motivation for using regularization, namely metric utility. We hypothesize that humans cannot perform the complex prewhitening of K_y^{-1}, which has long range voxel correlations, reflected in the complex structure of the weakly regularized template in Fig 3. Therefore, templates with more compact structure resembling the signal of interest could lead to more informative metrics relevant to human observer performance.

III. Method

A. Generation of Images

1) The Reconstructed Signal: The signal of interest was chosen to be an ellipsoid with axes of lengths 4 detector bin widths and 2 detector bin widths, and its location was fixed to the center of the field of view for reconstruction. We investigate a small high-contrast signal, which can be seen as a model of micro-calcification detection in breast CT. The signal was defined in the continuous object domain and discretized by a continuous-to-discrete forward projection operator, previously denoted as $\mathcal{P}$.

The reconstruction algorithm used was the BPF algorithm discussed in section II-A. Projection data were acquired over an angular range of 1.128π, corresponding to 180 degrees of trajectory around each chord used for reconstruction. 32 evenly spaced views were used with a 16×8 array of detector elements. The reconstruction was performed onto 128 chords with 32 points along each chord. The chords were arranged in three rows of 32 chords each, so that the reconstructed image volume was rectangular with $32 \times 32 \times 4$ voxels. The ratio of the source-to-detector distance to the source's radius of rotation was $8 : 5$. The reconstructed difference between signal hypotheses $\Delta\bar{y}$ is shown in Fig. 2. Visible artifacts in the reconstructed signal are a result of the discretization in the sinogram and image domains and the small size of the signal. In this study, we chose to use very coarse discretization so that the matrix K_y would have dimensionality low enough that it could be stored directly in computer memory and a direct matrix inverse could be obtained. However, it should be noted that this small system can be relevant to real CT applications, since this level of discretization is what might be seen in an actual microcalcification ROI.

2) Noise Model: After computing the discrete projection data $\mathbf{g} = \mathcal{P}f$, zero-mean, independent, identically distributed Gaussian noise was added to the projection data. The standard deviation of the additive noise was uniform across the detector elements and equal to roughly ten times the maximum value of the signal in the projection data domain. This implies that the data covariance vector K_g is diagonal such that

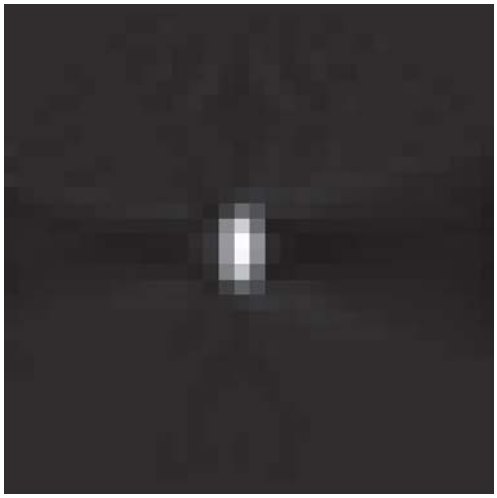


Fig. 2. The reconstructed noiseless signal $\Delta\bar{y}$, i.e., the mean difference between image hypotheses. The display window used is [-2,15].

$$(K_g)_{i,j} = \begin{cases} \alpha & : i = j \\ 0 & : i \neq j \end{cases} \qquad (11)$$

where α is a constant. We then have that the reconstructed image covariance matrix will be given by

$$K_y = AK_gA^T = \alpha AA^T \qquad (12)$$

where A is the reconstruction matrix described in Section II-A, and K_y is again the reconstructed image covariance matrix. Inspecting this expression for the image covariance matrix, it becomes obvious why the inversion of this matrix can be nontrivial. K_y has dimensionality $M \times M$, where M is the number of voxels in the reconstructed image. Further, various components of the matrix A such as the matrix representing the weighted back-projection step in the FBP algorithm make K_y non-diagonal.

B. Model Observer Computation

HO SNR was computed as in Eqn. 9, where the image covariance matrix K_y is given by Eqn. 12. The Hotelling SNR was computed by obtaining the singular value decomposition of K_y and constructing the Moore-Penrose pseudo-inverse to solve Eqn. 7. Solutions to the minimization problem (10) were obtained using the LSQR algorithm [6]. The damping parameter in (10) was varied from 10^{-3} to 10^2 with power of 10 increments. The resulting SNR of each regularized template was then compared to SNR_{Hot} via an efficiency metric defined by

$$\varepsilon = \frac{\text{SNR}^2}{\text{SNR}^2_{Hot}} \qquad (13)$$

Further, the number of iterations of the LSQR algorithm was stored in order to quantify the computational gains arising from obtaining a regularized solution.

IV. Results

Figure 3 shows the results of solving the optimization program (10) with varying degrees of regularization, as described above. Note that with increasing values of the damping parameter λ, the template recovers the structure of the signal, shown in Fig. 2. In fact, the template obtained with $\lambda = 100$ is practically indistinguishable from the signal.

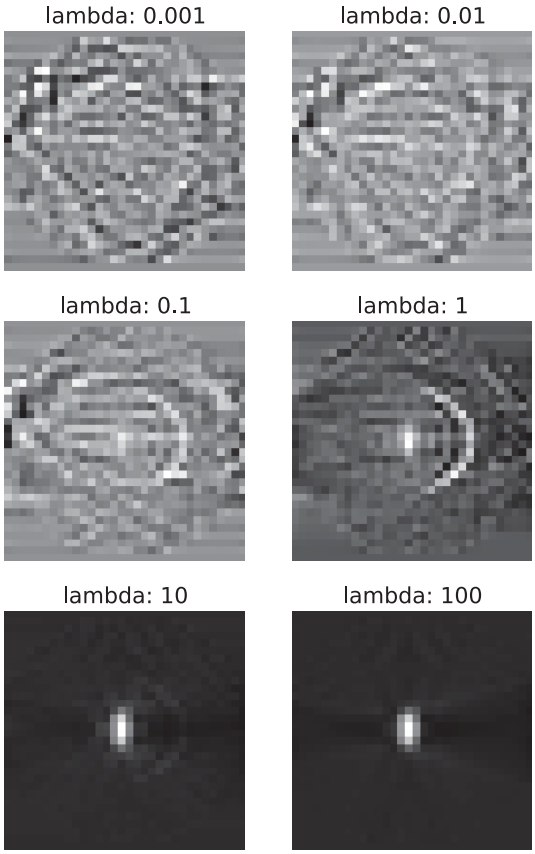


Fig. 3. Templates obtained through varying degrees of regularization of the solution to the Hotelling template equation, Eqn. 7. Note that the most heavily regularized template is practically indistinguishable from the mean signal $\Delta\bar{y}$ shown in Fig 2.

Table I presents the results of obtaining each of the six regularized templates shown in Fig. 3. Note that increasing the amount of regularization in the solution of the template greatly decreases the number of iterations needed to solve the optimization program (10).

TABLE I
Properties of Regularized Templates

λ	ε	# of iterations
0.001	0.99999	8192
0.01	0.99994	8192
0.1	0.99694	1579
1	0.95423	197
10	0.61955	25
100	0.47123	5

V. Conclusion

We computed SNR values for 6 templates that were computed as regularized versions of the 3D Hotelling template and compared these values to the Hotelling SNR. Further, we inspected the structures of the templates themselves (as the construction of the minimization program from which they are derived ensures their uniqueness), and noted the computational burden of calculating each template.

The point of regularization at which the templates begin to resemble the signal can be seen as the point at which the corresponding observer trivializes the prewhitening task in detecting the signal of interest. Given the intuitive view that humans are likely to perform detection tasks by a simple template matching paradigm, with a template that structurally resembles the signal, we hypothesize that it is approximately at this level of regularization that the model observer's performance becomes predictive of human observer performance. In this case, this would correspond to an efficiency of $\varepsilon \approx 0.5$.

If this hypothesis proves correct, there could be valuable implications regarding computational efficiency of these model observer metrics. Namely, it could be possible to obtain a metric which is predictive of human performance with a computational effort that is orders of magnitude lower than full computation of 3D Hotelling observer performance. In this case, a reduction of FLOPs by two orders of magnitude was observed, as seen in the third column of Table I. Future work could be performed to compare human observer performance through 2AFC trials to these model observer results in order to investigate a possible correlation between template regularization and a point at which these templates become predictive of human performance.

Acknowledgment

This work was supported in part by NIH R01 Grant Nos. CA120540, EB000225, and CA158446 and NIH T32 Grant No. EB002103-22. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2004.

[2] J. Beutel, H. L. Kundel, and R. L. V. Metter, *Handbook of Medcial Imaging*. SPIE Press, 2000, vol. 1 - Physics and Psychophysics.

[3] E. Y. Sidky, S. J. LaRoque, and X. Pan, "Accurate computation of the hotelling template for SKE/BKE detection tasks," in *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, ser. Proc. SPIE, B. Sahiner and D. J. Manning, Eds., vol. 6917, 2008, pp. 69 170W–1–69 170W–5.

[4] A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Bio.*, vol. 53, pp. 2471–2493, 2008.

[5] Y. Zou, X. C. Pan, and E. Y. Sidky, "Theory and algorithms for image reconstruction on chords and within regions of interest," *J. Opt. Soc. Am. A*, vol. 22, pp. 2372–2384, 2005.

[6] C. Paige and M. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.

Cone-beam Reconstruction and Evaluation for Helical Tomography

Guangzhi Cao[1], Brian Nett[1], Scott Hsieh[2], Jed Pack[3], and Jiang Hsieh[1]
[1]GE Healthcare, 3000 N. Grandview Blvd., W-1180, Waukesha, WI USA
[2]Department of Electrical Engineering, Stanford University, Stanford, CA USA
[3]GE Global Research Center, One Research Circle, Niskayuna, NY USA

***Abstract*—A variety of helical reconstruction approaches are investigated and evaluated for cone-beam helical scans including: the Feldkamp algorithm (FDK) without view weighting, a view weighted FDK algorithm, a theoretically exact Katsevich algorithm and an iterative reconstruction method. These approaches address the root causes of the cone-beam artifacts in helical reconstruction from different perspectives. Simulated wide cone-beam helical data is used to assess the relative merits of the different algorithms in terms of cone-beam artifacts, noise, and motion sensitivity.**

I. Introduction

As the detector size continues to increase in computed tomography (CT) helical scans, it becomes more difficult to reconstruct an image volume which is not corrupted by cone-beam artifacts. The cone-beam artifacts in helical scans are primarily due to truncation and mis-handled frequencies in weighting of redundant data. The truncation artifacts are caused by interrupted illumination where some of the voxels in the volume project off the detector for some views. The second problem which occurs is due to incorrect weighting of redundant measurements in Radon space (which can also be viewed as the local Fourier space of the image object). These errors due to incorrect frequency weighting typically manifest themselves as artifacts which have low in-plane spatial frequencies. For a more detailed description of these issues, please refer to another paper in this meeting [1].

Traditionally, the FDK-type algorithms [2] have been widely used by many of the major CT manufacturers because of their simplicity and the desirable and predictable image characteristics they produce. Within the general FDK framework, sophisticated view weighting methods are generally adopted to suppress the cone beam artifacts, while maintaining the desired noise statistics [3], [4]. Theoretically exact algorithms have been proposed for helical reconstruction by A. Katsevich and others [5]. By definition, these theoretically exact algorithms generate reconstruction image volumes with little or no cone-beam artifacts. However, they also lead to different noise characteristics compared to the FDK-type algorithms. More recently, iterative reconstruction (IR) algorithms have been proposed for CT reconstruction and showed great promise in reducing both noise and cone-beam artifacts [6], [7]. However, IR algorithms are generally associated with a significant computational penalty and therefore the reconstruction is slower than other methods. In this work, different approaches are investigated and evaluated for cone-beam helical reconstruction.

II. Methods

A. Notation

In a helical scan, the source trajectory can be expressed as

$$\underline{s}(\beta) = (R\sin\beta, R\cos\beta, \frac{H}{2\pi}\beta), \beta \in [\beta_s, \beta_e] \ , \qquad (1)$$

where R is the radius of the helical source trajectory, β_s and β_e correspond to the starting and ending view angles of the helical source trajectory, respectively, and H is the distance traveled by the source point per rotation along z-axis. The normalized helical pitch, denoted as h, is defined as the ratio of H over detector size at ISO. The cone-beam transform of a 3D object f (equivalently $f(x,y,z)$ or $f(\underline{x})$ with $\underline{x} = (x,y,z)$) can be written as

$$p(\underline{s},\underline{\theta}) := \mathcal{D}(f) = \int_0^{+\infty} f(\underline{s}(\beta) + t\underline{\theta})dt \ , \underline{\theta} \in S \ , \qquad (2)$$

where S is the set of all possible unit vectors in space. Note that the cone-beam projection is a function of both $\underline{\theta}$ and $\underline{s}$ parameterized by the angular position of the source point β. The objective of reconstruction is to obtain an estimate of the object f based on the measured projection data p.

It is common for reconstruction algorithms to implement the reconstruction in the so-called cone-parallel geometry that can be obtained through row-wise fan-to-parallel rebinning in the native cone-beam geometry [3]. In the rebinned cone-parallel geometry, any projection data p can be determined by the cone angle α, view angle β and orthogonal distance t from the ISO-ray. Therefore, we can re-parameterize the projection data $p(\underline{s},\underline{\theta})$ as $p_r(\alpha,\beta,t)$ in the cone-parallel geometry. Note that after row-wise cone-parallel rebinning, the curvature of the resultant virtual detector is inverted and the 1D ramp filtering is applied along the tangential direction of the helical source trajectory, which significantly improves reconstruction accuracy.

Next, we discuss various reconstruction algorithms that can be used for cone-beam helical scans.

B. FDK Algorithm Without View Weighting

In the cone-parallel geometry, one FDK-type algorithm to reconstruct an object $f(\underline{x})$ from data acquired with a helical scan without any view weighting can be written as

$$\hat{f}(\underline{x}) = \frac{\pi}{\beta_{\max}-\beta_{\min}} \int_{\beta_{\min}}^{\beta_{\max}} \tilde{p}_r(\alpha,\beta,t)\cos\alpha d\beta \ , \quad (3)$$

$$\tilde{p}_r(\alpha,\beta,t) = p_r(\alpha,\beta,t) \bigotimes g(t) \ , \quad (4)$$

where $g(t)$ denotes the 1D ramp filter kernel and $\bigotimes$ denotes the 1D convolution operator. The interval $[\beta_{\min},\beta_{\max}]$ defines the view range over which the projection data are used to reconstruct the image intersecting the helical trajectory at view angle $\beta = (\beta_{\min}+\beta_{\max})/2$. The view range is generally a multiple of π. The factor $\pi/(\beta_{\max}-\beta_{\min})$ normalizes the reconstruction when different view ranges are used. This cone-beam FDK-type algorithm is mathematically approximate in nature as it does not address either truncation or mis-handled frequencies. The FDK algorithm without view weighting can lead to significant cone beam artifacts and therefore is not commonly used for helical reconstruction in practice.

C. View Weighted FDK Algorithm

View weighting has been extensively employed as an improvement to the FDK algorithm [3], [4]. Here, we introduce a view weighting scheme that is an extension of the 3D view weighting approach proposed in [3] that has been demonstrated to work well when the detector size is less than or equal to 40mm at ISO center. Mathematically, it can be written as

$$\hat{f}(\underline{x}) = \int_{\beta_{\min}}^{\beta_{\max}} w(\alpha,\beta,t,r)\tilde{p}_r(\alpha,\beta,t)\cos\alpha d\beta \ , \quad (5)$$

where

$$w(\alpha,\beta,t,r) = \frac{w_{2d}(\beta,t)\tan^{k(h,r)}(\alpha_c)}{w_{2d}(\beta,t)\tan^{k(h,r)}(\alpha_c)+w_{2d}(\beta_c,t_c)\tan^{k(h,r)}(\alpha)} \ , \quad (6)$$

where $r := \sqrt{x^2+y^2}$ and the subscript c represents the corresponding parameters of the conjugate ray. Here w_{2d} can be any 2D view weighting function that is used traditionally in 2D reconstruction or in the case of a small cone angle. This 2D view weighting must meet the constraint

$$w_{2d}(\beta,t)+w_{2d}(\beta_c,t_c) = 1.0 \ . \quad (7)$$

The simplest case is a full scan where one may use $w_{2d}(\beta,t) = 0.5$. Note that the FDK algorithm without view weighting can also be viewed as constant view weighted FDK algorithm, i.e. $w(\alpha,\beta,t,r) = \pi/(\beta_{\max}-\beta_{\min})$.

In addition to its dependence on α, β and t as in its original form in [3], the new view weighting is also a function of the distance between the voxel to be reconstructed and the ISO, r. The additional dependence on r allows the view weighting function to apply different weighting strength on different voxels depending on their location. This is to take into account of the fact that the amount of measured projection data varies with respect to the location of a voxel. The view weighting function can be extended to various pitches through the dependence of k on the pitch h. Note that the specific value of $k(h,r)$ needs to be experimentally tuned in order to achieve optimal image quality (IQ) for different pitches.

By introducing the view weighting, the reconstruction can suppress most of the truncation artifacts by reducing the weights on the projection data that is not physically measured by the detector while using all the measured data to maintain a low noise level. However, the mis-handled frequency that is intrinsic to the cone-beam is not be well addressed by the view weighting.

D. Katsevich Algorithm

The Katsevich algorithm was proposed by A. Katseivch [5] and is theoretically exact for helical scans if the detector is large enough to capture a region slightly larger than the Tam-Danielsson window [8]It was developed based on the concept of π-line that connects the voxel to be reconstructed and two source points on the helical scan trajectory. Mathematically, the reconstruction in native cone-beam geometry can be written as follows

$$\hat{f}(\underline{x}) = -\frac{1}{2\pi^2}\int_{I_{\pi(\underline{x})}} \frac{1}{|\underline{x}-\underline{s}(\beta)|}\int_0^{2\pi}\frac{\partial}{\partial q}p(\underline{s}(q),\underline{\Theta}(\beta,\underline{x},\gamma))|_{q=\beta}\frac{d\gamma}{\sin\gamma}d\beta \ , \quad (8)$$

where $I_{\pi(\underline{x})}$ defines the π-line passing through $\underline{x}$, $\underline{\Theta}(\beta,\underline{x},\gamma) := \cos\gamma\underline{\theta}(\beta,\underline{x}) + \sin\gamma\underline{e}(\beta,\underline{x})$ and $\underline{e}(\beta,\underline{x}) := \underline{\theta}(\beta,\underline{x}) \times \underline{u}(\beta,\underline{x})$. For details, please see [5]. Note that the Katsevich algorithm is realized in a traditional filtered back projection framework and employs a shift-invariant filter, which makes the algorithm practical. The difference from traditional FDK algorithms is that the filtering is generally done along tilted lines instead of the detector row direction. In this way the measured frequencies for a given voxel can be handled correctly and therefore the reconstruction is exact if the projection data is not contaminated by noise. Our implementation of the Katsevich algorithm is based on native cone-beam geometry.

Although the Katsevich algorithm is mathematically accurate and elegant, it has thus far not been adopted by the vendors of diagnostic CT scanners. One major reason is that while the algorithm is exact, it is not optimal from the perspective of dose utilization due to its difficulty in handling redundant data. Another reason is the discontinuous nature of the view weighting in the Katsevich algorithm. These generally lead to non-uniform noise texture and motion artifacts as we will see in next section.

E. Iterative Reconstruction

Iterative reconstruction has been intensively studied as an alternative to analytical reconstruction algorithms in CT recently. There are two categories of iterative methods in CT reconstruction: one is statistical iterative reconstruction that generally performs a global minimization or maximization of a certain probability measure, such as maximum likelihood or maximum *a posteriori* probability [6]; another is more

straightforward effort of matching the measured data, which can be based on an algebraic solution [9] or iterations of FDK-type reconstruction [7]. Here we use the second class as an example that can be expressed as

$$\begin{aligned} &f_0 \leftarrow \mathcal{D}^{-1}(p) \\ &\text{For } k = 1, 2 \dots K \\ &\qquad f_k \leftarrow f_{k-1} + \lambda_1 \mathcal{D}^{-1}(p - \mathcal{D}(f_{k-1})) + \lambda_2 \mathcal{R}(f_{k-1}) \end{aligned} \tag{9}$$

Here the inverse operator $\mathcal{D}^{-1}(p)$ represents the reconstruction from the projection data p which, for example, can be an FDK algorithm; the forward re-projection operator $\mathcal{D}(f_k)$ simulates the line integral process of the helical scan based on the current reconstruction f_k, which can either be in native cone-beam geometry or rebinned cone-parallel geometry; and $\mathcal{R}(f_k)$ is a regularization term that is used to impose *a priori* knowledge, e.g. smoothness, on the reconstructed image. The iterative reconstruction handles the cone-beam artifacts (both truncation and mis-handled frequency) implicitly through enforcing fidelity with the measured data.

In this study, the view weighted FDK algorithm in cone-parallel geometry in Section II.C is used for the inverse operator and the distance driven re-projection in native cone-beam geometry is used for the forward operator [10]. Therefore, a rebinning operator is required after the re-projection although it is not explicilty shown in (9). No regularization term is used here. The iterative reconstruction is run for $K = 1$ iteration as most of the low frequency cone-beam artifacts are already effectively mitigated for the test cases considered here. Therefore, it is simply one more iteration of the view weighted FDK algorithm. With more iterations and more sophisticated design of the system and regularization models, it is possible to achieve higher resolution and lower noise variance; but this is not within the scope of this work.

III. Evaluation

In this section the reconstruction algorithms discussed above are evaluated using simulated cone-beam helical data from different IQ perspectives: cone-beam artifact, noise and motion sensitivity. The simulations were performed using the CatSim software. A cylindrical detector was simulated with a cone angle of 7.32 degree and a fan angle of 47.6 degree. The helical data are simulated with the following scan protocol: 70keV, 1.0 sec/rotation and a normalized pitch of 0.99. The helical body phantom (HBP) that was specially designed to challenge helical reconstruction algorithms and a more clinically relevant nCAT phantom are used in the simulations. Images were reconstructed using a 2π view range of data with a field of view of 500mm on an image grid of 512×512 and a slice spacing of 0.625mm.

A. Cone Beam Artifacts and Noise

To evaluate the cone-beam artifacts and the noise texture, a helical scan of the HBP phantom was simulated with a beam current of 300 mA. Reconstruction results of the different algorithms are shown in Fig. 1. It can be seen that the FDK reconstruction without view weighting is dominated by cone-beam artifacts (mostly truncation artifacts). The view weighted FDK reconstruction effectively mitigates most of the truncation artifacts. There is still some frequency mis-handling in the image which can be seen as dark shading near the air hole. The Katsevich reconstruction does not contain any low frequency cone-beam artifacts as expected. Compared to the view weighted FDK, the extra iteration enables the IR reconstruction to remove the residual cone-beam artifacts and improve the resolution which results in some aliasing due to the large pixel size used in this case.

Ignoring the cone-beam artifacts, the FDK without view weighting, view weighted FDK, and IR reconstructions show typical FBP noise texture while the Katsevich reconstruction shows high and non-uniform noise. The standard deviation measured in the rectangle ROI show are 29.7, 33.8, 73.9 and 37.3 HU in Fig. 1(a)-(d), respectively.

B. Motion Sensitivity

Motion exists in clinical helical scans that can be caused by undesired patient respiratory and cardiac motion. Therefore, it is important for reconstruction algorithms to be robust to motion of at least a certain level. To evaluate this, a helical scan of a dynamic HBP phantom was simulated with a beam current of 300 mA, where the HBP was moving from left to right with a speed of 1 mm/sec. Reconstruction results of the different algorithms are shown in Fig. 2. From the results it can be seen that the FDK-type algorithms are more robust to this level of motion while motion artifacts in the form of streaks and object shifts are clearly visible in the Katsevich and IR reconstructions, respectively.

C. Overall Image Quality

To evaluate overall IQ in a more clinically relevant setting, a helical scan of the nCAT phantom was simulated with a beam current of 500 mA. The axial images and the reformatted images of different reconstructions are shown in Fig. 3. In general the results agree well with what was observed in the HBP simulations, i.e., the FDK reconstruction without view weighting has overwhelming cone-beam artifacts mostly due to truncation; the view weighted FDK reconstruction contains slight cone-beam artifacts due to mis-handled frequency (more visible in the reformatted image); the Katsevich reconstruction has high and non-uniform noise but with little cone-beam artifacts; and the IR reconstruction reduces the residual cone-beam artifacts while having higher resolution and noise as compared to the view weighted FDK reconstruction.

IV. Conclusion

A comprehensive investigation and evaluation of reconstruction approaches for wide cone-beam helical scans has been provided. While the cone angle of the beam is too large to be ignored, the cone-beam artifacts can be effectively mitigated through view weighting, exact reconstruction or iterative reconstruction. However, different algorithms demonstrated different image properties. An appropriate reconstruction approach should be chosen in practice depending on what is

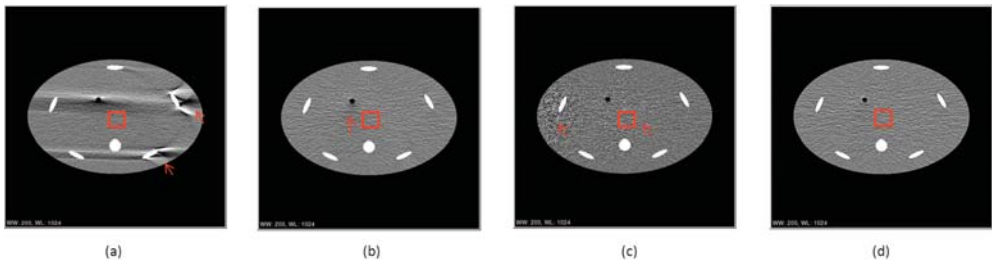


Fig. 1. Cone-beam artifacts and noise comparison of different reconstruction algorithms for simulated HBP cone-beam helical scan: (a) FDK without view weighting; (b) View weighted FDK; (c) Katsevich algorithm; (d) IR. The arrows in (a) and (b) point to cone-beam artifacts. The arrows in (c) point to non-uniform noise. The standard deviations in the rectangle ROIs are 29.7, 33.8, 73.9 and 37.3 in (a), (b), (c) and (d), respectively. (W/L = 200/0)

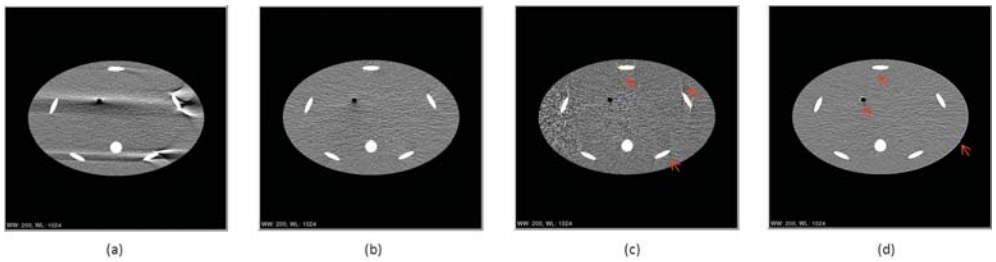


Fig. 2. Motion sensitivity comparison of different reconstruction algorithms for simulated dynamic HBP cone-beam helical scan: (a) FDK without view weighting; (b) View weighted FDK; (c) Katsevich algorithm; (d) IR. The arrows point to motion artifacts. (W/L = 200/0)

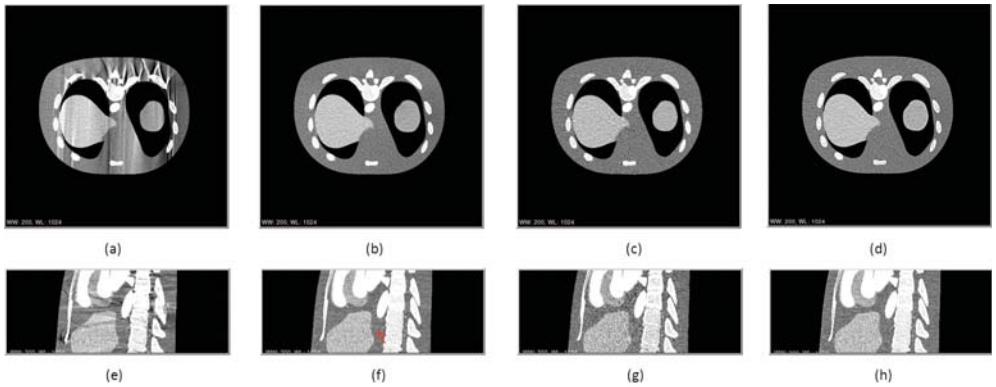


Fig. 3. Overall image quality comparison of different reconstruction algorithms for simulated nCAT cone-beam helical scan: (a)(e) FDK without view weighting; (b)(f) View weighted FDK; (c)(g) Katsevich algorithm; (d)(h) IR. The arrow points to cone-beam artifacts caused by mis-handled frequency. (W/L = 200/0)

the most relevant factor as well as other practical constraints (e.g. temporal resolution) for a given application.

REFERENCES

[1] J. Pack, Z. Yin, K. Zeng, and N. B, "Mitigating cone-beam artifacts in cardiac CT imaging on a large cone-angle system," in *Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2013, p. Submitted.

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, p. 612, 1984.

[3] X. Tang, J. Hsieh, R. Nilsen, S. Dutta, D. Samsonov, and A. Hagiwara, "A three-dimensional-weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric cthelical scanning," *Physics in Medicine and Biology*, vol. 51, no. 4, p. 855, 2006.

[4] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP - a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch," *Physics in medicine and biology*, vol. 49, no. 11, p. 2209, 2004.

[5] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral ct," *SIAM Journal on Applied Mathematics*, vol. 62, no. 6, pp. 2012–2026, 2002.

[6] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[7] J. Sunnegårdh and P. Danielsson, "Regularized iterative weighted filtered backprojection for helical cone-beam CT," *Medical physics*, vol. 35, p. 4173, 2008.

[8] K. Tam, S. Samarasekera, and F. Sauer, "Exact cone beam ct with a spiral scan," *Physics in medicine and biology*, vol. 43, no. 4, p. 1015, 1999.

[9] A. Andersen and A. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[10] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in medicine and biology*, vol. 49, no. 11, p. 2463, 2004.

Penalized weighted least-squares image reconstruction for low-dose CT using adaptive MRF coefficients predicted from normal-dose scan

Hao Zhang, Hao Han, Yan Liu, Hongbing Lu, Jianhua Ma, Jing Wang, and Zhengrong Liang

***Abstract*–Low-dose computed tomography (LDCT) is desirable due to the growing concerns about excessive radiation exposure in clinic. One common strategy to achieve low-dose is to lower the mAs setting in currently available CT scanners, but the image reconstructed by conventional filtered backprojection (FBP) is severely degraded due to excessive quantum noise. Statistical reconstruction (SR) algorithms have shown the potential to significantly improve the reconstructed image quality from the noisy projection. In this work, we present a penalized weighted least-squares (PWLS) iterative reconstruction scheme to improve low-mAs CT reconstruction by incorporating a quadratic-form Gaussian Markov random field (MRF) regularization with coefficients adaptively predicted from the corresponding normal-dose scan. More specifically, given the FBP reconstruction result of the normal-dose scan, we compute its object scale map (which describes the uniformity of local structures) and exploit the map to determine the MRF window size in a pixel-by-pixel fashion. The rationale behind this is that the more uniform the local structure is, the smaller region the frequency spectrum will spread, so a smaller MRF window should be adopted, and *vice versa*. For each pixel, we adopt a reasonable larger sample window to predict the MRF coefficients via least-squares regression. The MRF coefficients, adaptively predicted from the normal-dose scan, are then used to help the following up low-dose scans reconstruction. Experimental results with both physical phantom and patient data demonstrated that the presented regularization is superior to conventional quadratic-form Gaussian MRF regularization using a fixed window and equal coefficients for neighbors of equal distance. Further quantitative evaluations of the proposed method are still under progress.**

***Index Terms* – low-dose CT, penalized weighted least-squares, Markov random field, predicted coefficients, normal-dose scan**

I. Introduction

Low-dose CT has attracted increasing attentions due to growing concerns about radiation exposure and the potential side effects in clinical practice. One simple and cost-effective way to reduce the radiation dose is to lower the X-ray tube mAs setting in currently available CT scanners when acquiring projection data. However, the quality of reconstructed CT images by FBP from the low-mAs acquisition will be degraded dramatically due to the excessive noise.

Many methods [1-11] have been proposed to improve the reconstruction of low-dose CT from noisy projections. Although the filter based algorithms [1-3] are computationally efficient and can suppress the noise to some extent, they often sacrifice structural details which could be critical in clinic, due to a lack of noise modeling. Statistical reconstruction (SR) algorithms [4-11], which take into account of statistical noise properties with accurate modeling, are superior to FBP in suppressing quantum noise and artifacts. SR algorithms reconstruct the CT images by maximizing or minimizing a cost function for an optimal solution in a statistical sense, where the cost function usually consists two terms. The data-fidelity term models the statistics of measured data, and the regularization term reflects *a prior* information. One commonly used regularization term is the quadratic-form Gaussian Markov random field (MRF) [5][7], which uses equal coefficients for neighbors of equal distance without considering discontinuities in images, thus may lead to over smoothing of edges or fine structures. Some non-quadratic penalties can preserve edges, but they often introduce piece-wise constant blocky artifacts and the results are also sensitive to the hyper-parameter that controls the shape of the penalty function [12]. In addition, they make the cost function non-quadratic and may complicate the computation.

Recently, Wang et al. [13] explored the potential of larger neighborhoods in MRF to include higher frequencies for the *prior* model. They utilized a large MRF window and estimated the MRF coefficients from a high-quality bone-region sample image, and incorporated this single set of MRF coefficients into the iterative CT reconstruction framework. Their preliminary results demonstrated the inner bone emphasis similar to that of conventional FBP with commercial bone kernel. Inspired by their work, we investigate to design a quadratic-form Gaussian MRF regularization term which exploits space-variant MRF window size and adaptive coefficients predicted from local sample window of a normal-dose scan, to help the following up low-dose scans CT reconstruction. In this way, both the MRF window size and corresponding coefficients are adaptive and could better reflect the *prior* information in space and frequency domain. Thus, it is expected that with the proposed regularization term predicted from a normal-dose CT scan, the following up low-dose scans reconstruction could be substantially improved.

This work was supported in part by the NIH/NCI under Grant #CA143111 and #CA082402. *Asterisk indicates corresponding author.*

H. Zhang and *Z. Liang are with the Departments of Radiology and Biomedical Engineering, Stony Brook University, NY11794, USA.

H. Han and Y. Liu are with the Department of Radiology, Stony Brook University, NY11794, USA.

H. Lu is with the Department of Biomedical Engineering, Fourth Military Medical University, Shanxi710032, China.

J. Ma is with the School of Biomedical Engineering, Southern Medical University, Guangdong510515, China.

J. Wang is with the Department of Radiation Oncology, University of Texas Southwestern Medical Center, TX 75390, USA.

II. METHODS

A. Statistical model

With the assumption of monochromatic x-ray generation, the acquired CT transmission data I can be described by:

$$I_i = \text{Poisson}(\lambda_i) + \text{Normal}(m_{e,i}, \sigma^2_{e,i}) \quad (1)$$

where λ_i is the mean number of x-ray photons collected by the detector bin i in repeated measurements, $m_{e,i}$ and $\sigma^2_{e,i}$ are the mean and variance of the electronic noise associated with the measurement I_i, respectively. For practice applications, $m_{e,i}$ is often calibrated to be zero. Based on this noise model, a formula of relationship between the mean and variance of the log-transformed projection data was proposed in [14]:

$$\sigma^2_{p_i} = \frac{1}{I_0^i}\exp(\bar{p}_i)\left(1+\frac{\sigma^2_{e,i}-1.25}{I_0^i}\exp(\bar{p}_i)\right) \quad (2)$$

where p_i is the log-transformed projection datum (which is also called the line integral) in detector bin i and $\bar{p}_i$ is the mean value of it in repeated measurements, I_0^i is the mean number of incident photons along the projection path i.

B. PWLS iterative image reconstruction

Based on the noise properties of CT projection data, the PWLS cost function in the image domain can be written as [7]:

$$\Phi(\mu) = (p - A\mu)^T \sum^{-1} (p - A\mu) + \beta R(\mu) \quad (3)$$

The first term in (3) is a weighted least-squares (WLS) measure, where p is the vector of the log-transformed projection data, μ is the vector of attenuation coefficients to be reconstructed, A is the system matrix with its element A_{ij} the intersection length of projection ray i with pixel j, $\sum$ is a diagonal matrix with its ith element the variance of line integral at the detector bin i estimated from (2). The symbol T and -1 herein is transpose and inverse operator respectively. The second term in (3) is a roughness penalty or a regularization constraint, where the smoothing parameter β controls the strength of regularization.

C. Regularization terms

Under the Gaussian MRF model, a quadratic-form penalty is widely used for iterative image reconstruction:

$$R(\mu) = \sum_j \sum_{m \in N_j} b_{jm}(\mu_j - \mu_m)^2 \quad (4)$$

where index j runs over all the pixels in the image domain, N_j denotes the small fixed neighborhood (typically 8 neighbors in the 2D case while 26 in the 3D case) of the jth image pixel, and b_{jm} are the directional weighting coefficients. The coefficients were usually set 1 for first-order neighbors and $1/\sqrt{2}$ for second-order neighbors. The major drawback of this regularization is that it only considers distance information but fails to take the discontinuities in images into account, thus may lead to oversmoothing of edges or fine structures.

Inspired by the work of Wang *et al* [13], we investigate to design a quadratic-form Gaussian MRF regularization term which exploits space-variant MRF window size and adaptive coefficients predicted from local sample window of a normal-dose scan, to help the following up low-dose scans CT reconstruction.

$$R(\mu) = \sum_j \sum_{m \in N_j} b_{jm}^{ND\_predict}(\mu_j - \mu_m)^2 \quad (5)$$

however, N_j is no longer a fixed neighborhood, but varies based on the object scale. Also, the coefficients are no longer constant but adaptively predicted from the normal-dose scan.

1) Computation of object scale

The scale at every pixel j, $K(j)$, is defined as the radius of the largest hyperball centered at the pixel j such that all pixels within the ball satisfied a predefined image intensity homogeneity criterion [15]. It is computed as follows: $K(j)$ is initialized as one. Then we iteratively increase the ball radius r by one and check the value of $FO_r(j)$ ("fraction of object"). When the first time this fraction falls below t_s (usually set to be 85%), we consider that the ball contains an object region different from that to which j belongs and set the scale value of $K(j)$ as $r - 1$ [16].

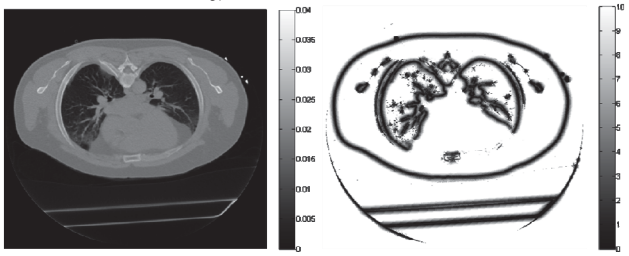


Fig. 1: (a) One transverse slice reconstructed by FBP with hanning filter from the sinogram data acquired with 100mAs (b) corresponding object scale map.

2) Determination of MRF/sample window size

For each pixel in the FBP reconstructed image from the normal-dose scan, we have an specific MRF window for it with the size determined by the object scale, which is a quantitative measure of local uniformity described above. The rationale is that the more uniform(larger object scale) a local region is, the smaller area the frequency spectrum will spread, so a lower-order MRF penalizing only differences among immediately neighboring pixels is enough. On the contrary, the more non-uniform a local region is, the larger area the frequency spectrum will spread, so a higher-order MRF should be adopted. Then for each pixel's MRF coefficients, they were predicted adaptively from a local sample window.

Statistically, to ensure the power of a least-squares linear regression with a certain number of predictors, the required sample size need to be determined. For our situation, to achieve a prediction power of at least 90% at the medium effect size (Cohen's $f^2 = 0.15$) given a significance level of less than 5%, the sample window size can be calculated by using the G*Power software [17], and the results are summarized in the following table.

Table1: MRF and sample window setting based on object scale

Object scale	MRF window size	Num.of predicted coefficients	Sample needed	Sample window size
>7	*3×3*	*8*	*136*	*13×13*
6-7	*5×5*	*24*	*206*	*15×15*
4-5	*7×7*	*48*	*278*	*17×17*

2-3	9×9	80	355	19×19
0-1	11×11	120	435	21×21

Then, for each pixel in the image, we have an adaptive MRF window size and corresponding sample window. Figure 2 shows us an example of that for a pixel in the image. Similar MRF/sample window choosing is performed for pixels in the image one by one.

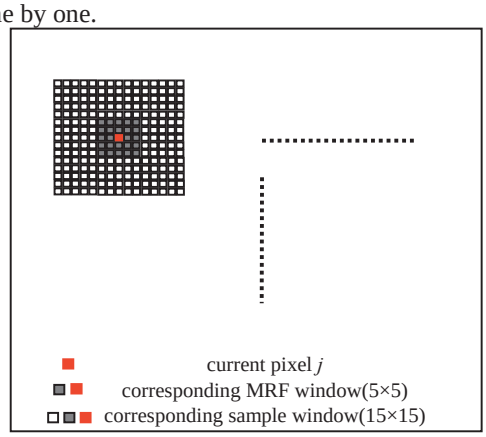


Fig. 2: Illustration of MRF and sample window choosing for one pixel.

3) Prediction of MRF coefficients

After we have the MRF and sample window size for each pixel in the FBP reconstructed image of the normal-dose CT scan, we can predict the corresponding MRF coefficients for each pixel via least-squares regression, because every pixel could be predicted from a linear combination of its clique-mates. The clique-mates of current pixel here are the neighbors within the MRF window. According to [13], the set of coefficients for the clique-mates can be formulated as:

$$\mathbf{b}_j^{ND\_predict} = \arg\min_{\mathbf{b}_j} \sum_{k \in S_j} (\mu_k^{ND} - \mathbf{b}_j{}^T \boldsymbol{\mu}_{N_k}^{ND})^2 = \left[\sum_{k \in S_j} (\boldsymbol{\mu}_{N_k}^{ND} (\boldsymbol{\mu}_{N_k}^{ND})^T) \right]^{-1} \left[\sum_{k \in S_j} (\boldsymbol{\mu}_{N_k}^{ND} \mu_k^{ND}) \right] \quad (6)$$

where $\mathbf{b}_j$ is the vector of MRF coefficients for the neighbors of pixel j within the MRF window, S_j is the sample window of pixel j, N_k represents the set of neighbors for pixel k within the MRF window and $\boldsymbol{\mu}_{N_k}$ is the vector of corresponding attenuation value for them. The symbol *ND* is short for normal-dose.

D. Summary of presented PWLS reconstruction method

With the weighted least-squares (WLS) criterion and the new designed quadratic-form Gaussian MRF regularization presented above, our PWLS iterative reconstruction scheme for low-dose CT reconstruction can be summarized as follows:

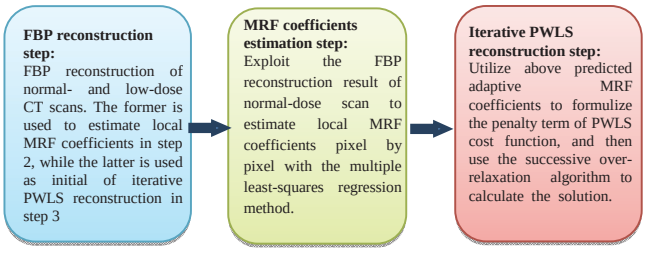


Fig. 3: Flow chart of proposed PWLS reconstruction method for low-dose CT.

$$\mu^* = \arg\min_{\mu \geq 0} \{ (p - A\mu)^T \Sigma^{-1} (p - A\mu) + \sum_j \sum_{m \in N_j} b_{jm}^{ND\_predict} (\mu_j - \mu_m)^2 \} \quad (7)$$

The cost function is of quadratic form and the iterative successive over-relaxation (SOR) algorithm could be used to calculate the solution. Details about the implementation are described in the appendix of [10].

III. Results

A. Anthropomorphic head phantom study

An anthropomorphic head phantom was used to evaluate the performance of the presented algorithm. The projection data were acquired by ExactArms of a Trilogy(tm) treatment system. The tube current was set at 80mA(normal dose) and 10mA (low dose) respectively, with X-ray pulse duration at each projection view 10 ms. One slice of the reconstructed images are shown in Fig. 4. The noise level (red square labeled 40×40 uniform region, std=7.5×10^{-4} mm^{-1}) of the reconstructed images in (c) and (d) match that of the 80mA image shown in (a).

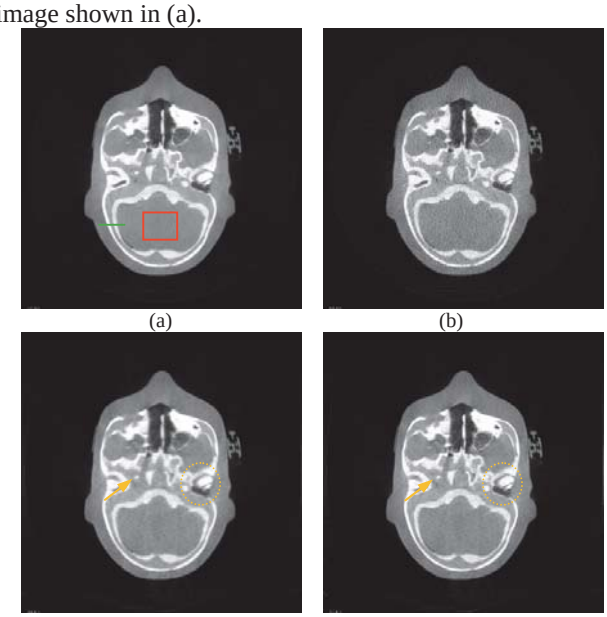


Fig.4: Reconstructed transverse slice of the head phantom:(a) FBP reconstruction from 80mA sinogram. (b) FBP reconstruction from the 10mA sinogram. (c) PWLS-isotropic reconstruction from the 10mA sinogram. (d) Proposed reconstruction from the 10mA sinogram. The display window is [0 0.03] mm^{-1}.

B. Patient data study

The projection data of a lung patient was acquired using a Siemens SOMATOM Sensation16 CT scanner. The tube current was set to 200mA and rotation speed is 0.5 seconds per rotation, that was 100mAs. We regarded this acquisition as the normal-dose scan, and instead of scanning the patient twice, we simulated the corresponding low-dose projection by adding noise to the normal-dose projection data based on equation (1) with I_0=5e4 and $\sigma^2_{e,i}$ =10. Each rotation included 1160 projection views evenly spanned on a circular orbit, and each view included 672 detector bins. One slice of the reconstructed images are shown in Fig. 5. The noise level (red square labeled 40×40 uniform region, std=1×10^{-3} mm^{-1})

of the reconstructed images in (c) and (d) match that of the normal-dose scan image shown in (a).

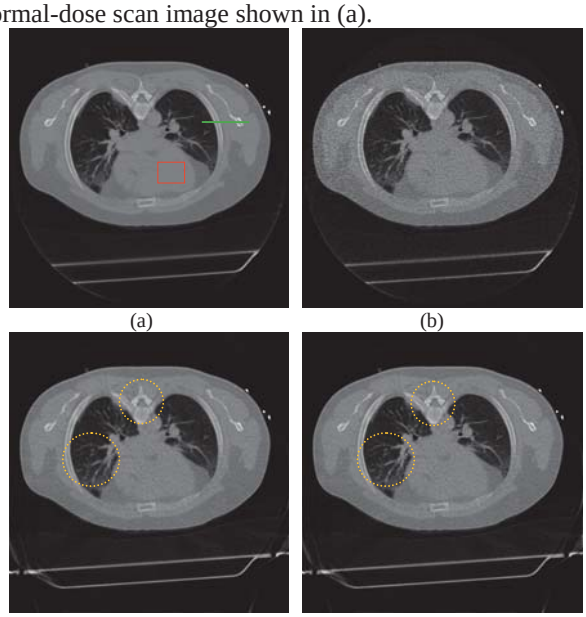


Fig.5: Reconstructed transverse slice of the patient: (a) FBP reconstruction from the normal-dose sinogram. (b) FBP reconstruction from the simulated low-dose sinogram. (c) PWLS-isotropic reconstruction from the simulated low-dose sinogram. (d) Proposed reconstruction from the simulated low-dose sinogram. The display window is [0 0.04] mm^{-1}.

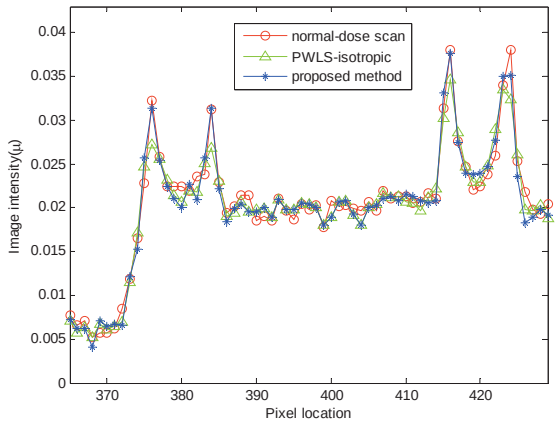


Fig. 6: Horizontal profiles located at the pixel position y=172 labeled in Fig 5(a). The corresponding methods are illustrated in figure legend.

IV. Conclusion and Discussion

In this work, we presented a penalized weighted least-squares (PWLS) iterative reconstruction method for low-mAs CT reconstruction by incorporating a quadratic-form Gaussian MRF regularization with adaptive MRF window size and adaptive coefficients predicted from the adaptive local sample window in the FBP reconstructed image of the corresponding normal-dose scan. Preliminary reconstruction results showed advantages over that with conventional quadratic-form Gaussian MRF penalty. However, in this proof-of-concept study, the image registration between the low-dose scan and corresponding normal-dose scan was not fully considered temporarily. This issue will be detailedly discussed in our future work. Also, further quantitative evaluations of the proposed method using more data are needed to demonstrate the advantages of proposed method.

References

[1] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Med. Phys.,* **25**, 2139-2147(1998).

[2] M. Kachelriess, O. Watzke, and W. Kalender, "Generalized multidimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT," *Med. Phys*., **28**, 475–490 (2001).

[3] C. Moore, T. Marchant, and A. Amer, "Cone beam CT with zonal filters for simultaneous dose reduction, improved target contrast and automated set-up in radiotherapy," *Phys. Med. Biol.,* **51**, 2191–2204 (2006).

[4] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Process*, **41**, 534–548 (1993).

[5] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Process*, **5**, 480–492 (1996).

[6] L. Elbakri and J. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography," *Proc. SPIE Medical Imaging,* **5032**, 1839–1850 (2003).

[7] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.,* **25**, 1272-1283(2006).

[8] J. Wang, T. Li, and L. Xing, "Iterative image reconstruction for CBCT using edge-preserving prior," *Med. Phys.,* **36**, 252-260(2009).

[9] J. Tang, B. Nett, G. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Phys. Med. Biol.,* **54**, 5781–5804 (2009).

[10] Q. Xu, H. Yu, X. Mou, L. Zheng, J. Hsieh, G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imaging,* vol. **31**, pp. 1682-1697(1998).

[11] M. Beister, D. Kolditz, W. Kalender, "Iterative reconstruction methods in X-ray CT," *Physica Medica*, **28,** 94-108(2012)

[12] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imag.,* **31**, 2194-2204(2012).

[13] J. Wang, K. Sauer, J. Thibault, Z. Yu and C. Bouman, "Prediction coefficients estimation in Markov random field for iterative X-ray CT reconstruction", *Proc.SPIE Medical Imaging*, **8314** (2012).

[14] J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen and H. Lu, "Variance analysis of x-ray CT sinograms in the presence of electronic noise background," *Med. Phys.,* **39,** 4051–4065(2012).

[15] P. Saha, and J. Udupa, "Scale-based image filtering preserving boundary sharpness and fine structure," *IEEE Trans. Med. Imag.*, **20**, 1140-1155 (2001).

[16] J. Liu, J. Yao, and R. Summers, "Scale-based correction for computer-aided polyp detection in CT colonography," *Med. Phys.,* **35**, 5664-5671 (2008).

[17] F. Faul, E. Erdfelder, A. Buchner, and A. Lang, "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses," *Behavior Research Methods*, **41**, 1149-1160(2009).

Dual-energy performance of x-ray CT with an energy-resolved photon-counting detector in comparison to x-ray CT with dual kVp

Yannan Jin, Geng Fu, Vladimir Lobastov, Peter M. Edic, Bruno De Man

***Abstract*—The dual-energy performance of two x-ray CT configurations was investigated: CT with fast kVp switching and with an energy-discriminating photon-counting detector. The pile-up effect in a photon-counting detector was simulated with a pulse pile-up model. Both the signal-to-noise ratio (SNR) of a monochromatic image and the decomposition error were used as the evaluation criteria. A range of performance was considered for the photon-counting detector capability. The impact of the pile-up effect was dependent on the detector's maximum periodic counting rate (N_0). When N_0 was 1 Mcps/mm^2, the pile-up effect corrupted almost all simulation cases. For N_0 equaling 10 Mcps/mm^2, about half of the simulation cases were adversely affected. When N_0>100 Mcps/mm^2, the pile-up effect did not have significant impact on performance metrics except for small phantoms with high mA. Fast kVp switching methods perform favorably when compared to the current state of the art systems utilizing photon-counting detectors. For an ideal photon-counting CT system (without pile-up effect), the entitlement is about 5%-15% higher than that of fast kVp switching in terms of SNR in monochromatic images. The count rate performance and clinical scan conditions are essential in evaluating the potential of a photon-counting system for routine clinical imaging.**

***Index Terms*—dual-energy, fast kVp switching, photon-counting detector, pulse pile-up**

I. INTRODUCTION

DUAL-energy computed tomography (DECT) differentiates and classifies materials by utilizing attenuation values acquired using two different energy spectra. The idea of DECT dates back to the early 1970s when Hounsfield first proposed to use an image-based method to distinguish iodine and calcium [1]. Alvarez and Macovski first proposed the projection-based material decomposition method, which decomposed measured attenuation coefficients using two linear basis functions, basically corresponding to the photoelectric absorption and the Compton Effect [2]. The relative amounts of two materials were obtained after decomposition, from which monochromatic images at different energy levels could be calculated.

Yannan Jin, Peter M. Edic and Bruno De Man are with the CT Systems and Applications Laboratory, GE Global Research, Niskayuna, NY 12309, USA. Email: jin@ge.com, edic@ge.com, deman@ge.com.
Geng Fu and Vladimir Lobastov are with the Detector Laboratory, GE Global Research, Niskayuna, NY 12309, USA. Email: gfu@ge.com, lobastov@ge.com

Several CT configurations have been developed to realize the dual-energy technique. One of the first configurations used multi-layer detectors [3], an approach which suffers from poor spectral separation. Dual-source technology [5] and the rotate-rotate acquisition approach resolve this challenge, but are highly sensitive to object/patient motion, due to the temporal delay between the high- and the low-kVp measurements. In this paper we focus on the two CT configurations that combine good spectral separation with low or no temporal delay: kVp switching [4] and photon-counting detectors [6]. Rapid tube voltage switching was first attempted about two decades ago [4], but only recently became practical thanks to advances in x-ray generation and detection technologies. In addition to offering energy-resolved imaging, photon-counting detectors have the additional advantages that they significantly reduce the impact of electronic noise and swank noise [6-10].

In spite of the inspiring future of photon-counting detectors, there are still technical challenges impeding its practical use in clinical applications. One of the main obstacles is the restriction on maximum detectable flux rate. Clinical CT has photon fluxes of approximately 10^9 photons/s•mm^2 in regions with minimal attenuation between the x-ray source and detector, which is well beyond the limits of the photon-counting detectors available today [10]. In this study we simulated a realistic photon-counting detector model including pulse pile-up, using a model proposed by Adam Wang, *et al*. in 2011 [11].

The aim of this study is to compare the dual-energy performance between x-ray CT systems utilizing photon-counting detectors and x-ray CT systems using fast kVp switching techniques. For both configurations, the dual-energy performances were quantitatively evaluated for different phantom types/sizes and different mA levels. The pile-up effect of the photon-counting detector was simulated and evaluated. Signal-to-noise ratio in monochromatic images and

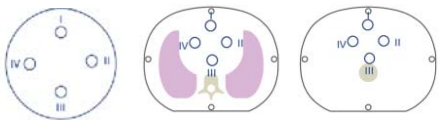


Figure 1: Phantoms used in the simulation study: (a) circular phantom, (b) thorax phantom and (c) abdomen phantom. The diameters of small, medium and large circular phantoms correspond to 16 cm, 24 cm and 32 cm respectively. The cross sections of small, medium and large thorax and abdomen phantoms correspond to 30 cm x 22 cm, 32.5 cm x 25 cm, 39 cm x 31 cm (width x height), respectively.

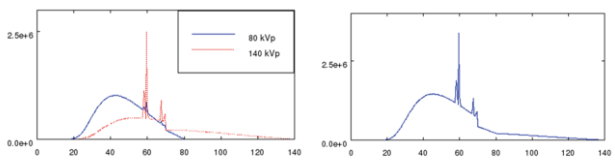


Figure 2: Spectra for the simulation study of kVp switching (left) and photon counting CT (right). The spectrum for photon counting CT is synthesized by summing the two spectra for kVp switching, which were weighted according to the ratio of tube currents at 80 kVp and 140 kVp.

material decomposition accuracy were defined as the figures of merit used for comparison.

II. MATERIALS AND METHODS

A. Scanning modality

In this study, we used the geometry of a 64-row clinical CT scanner (Discovery CT750 HD, GE Healthcare, Waukesha, WI) and a fast switching protocol of 80/140kVp. For photon-counting CT, the same scanning geometry was used and the detector module was simulated as an energy-discriminating photon-counting detector providing two energy bins.

B. Phantoms

Both circular water phantoms and semi-anthropomorphic CIRS phantoms (CIRS, Norfolk, VA) were used in simulations, as shown in Figure 1. Each phantom has three different sizes corresponding to small, medium and large objects. The four inserts in each phantom were iodine rods with concentrations of 1, 2, 5, 10 mg of iodine per mL.

C. Spectrum

The spectra for kVp switching and photon counting (PC) CT system configurations are shown in Figure 2. The spectrum for photon-counting CT was synthesized by adding the two spectra for kVp switching, which were weighted according to the ratio of tube currents. This setting ensures the match of dose level in photon-counting and kVp switching configurations, which simplifies the evaluation process.

D. Photon-counting detector

A dedicated software module was developed and added to the CT simulation environment called CATSim [12]. The module works as a callback function to simulate the pulse pile-up effect in a photon-counting detector. The physical effects that are simulated are:

- Detection efficiency
- Pile-up distortion [11]
- Energy binning
- Data Acquisition System (DAS)
- Electronic noise

In the simulation of photon-counting CT including detector pile-up effects, the highest photon flux occurs in an "air" acquisition - all the photons arrive at the detector without being attenuated by the object. In this study we used the air scans without pile-up effect, assuming that pile-up in air scans can be eliminated by scaling the air scan acquired at lower mA (and averaged to improve signal-to-noise ratio).

E. Evaluation criteria

The contrast of a certain material was determined by averaging the CT values in a homogenous region of interest (ROI). Quantum noise was added to simulated projection data. After reconstruction, the image noise level was determined by calculating the standard deviation of the CT values in the central ROI.

To evaluate the dual-energy performance, we focused on the monochromatic image which has two figures-of-merit (FOM). The first FOM is the signal-to-noise ratio (SNR) which is defined as

$$SNR_{Mono} = \frac{Contrast_{Mono}}{Noise_{Mono}}.$$

The second FOM is the decomposition error in the monochromatic image, which is defined as

$$Error_{Mono} = \left|\frac{Contrast_{Mono} - Contrast_{Iod}}{Contrast_{Iod}}\right| \times 100\%$$

Where $Contrast_{Mono}$ is the averaged CT value of the iodine insert in the monochromatic image and $Contrast_{Iod}$ is the „true" value of iodine at 70 keV:

$$Contrast_{Iod} = \frac{\mu_{Iod}(70\ keV) - \mu_{Water}(70\ keV)}{\mu_{Water}(70\ keV)} \times 1000$$

F. Simulation scheme

The simulation of both kVp switching and photon-counting (with and without pile-up effect) CT configurations was conducted using spectra generated with tube voltages of 80 and 140 kVp. Baseline phantom configurations were scaled for both the thorax and abdomen phantoms to emulate lean, normal and obese patients. The SNR of monochromatic images was evaluated. The energy threshold of photon-counting CT was determined by matching the noise level in the low- and high-energy bins.

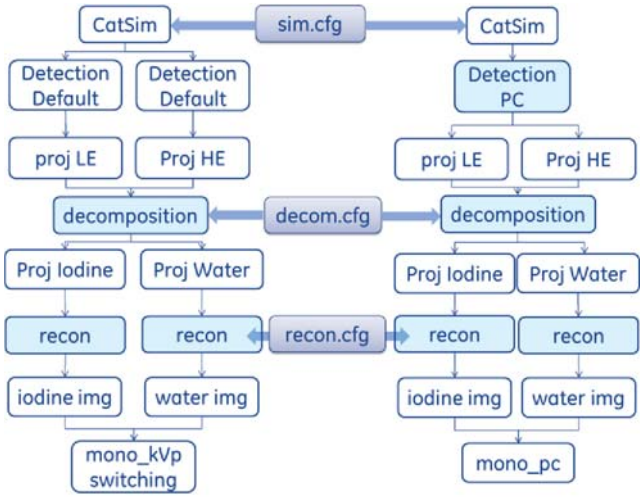


Figure 3 Flowchart for projection data generation, material decomposition, and the generation of monochromatic images for both kVp switching (left) and photon-counting CT (right).

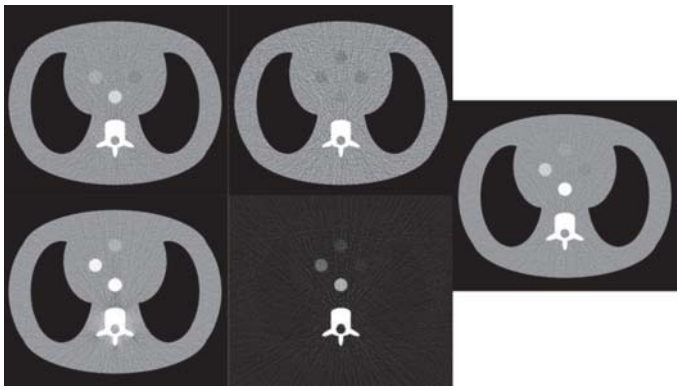


Figure 4 Decomposition results of the medium thorax phantom with kVp switching. The left column corresponds to the reconstructed images at low (bottom) and high energy (top); the central column is the density distribution of water- (top) and iodine-equivalent (bottom) images; the right column is the synthesized monochromatic image.

The flowchart of the simulation software is shown in Figure 3. The CATSim software was executed with different detection modules to generate requisite projection data. Note that kVp switching and photon-counting CT configurations shared the same configuration file to ensure the same scanning geometry and relevant scanning parameters.

After the projection data were generated, material decomposition (MD) was conducted to decompose projection data into water- and iodine-density integrals, which were subsequently reconstructed into density images and then combined as monochromatic images. Note that the MD and reconstruction also shared the same configuration file to ensure the same setting for dual-kVp and photon-counting CT configurations.

III. Results

A. kVp switching vs. photon counting

The simulation results in Figure 4 illustrate results of the material decomposition process. The x-ray tube was operated at 80/140 kVp and tube currents were scaled appropriately to match the projection noise. The projection data at low and high energy were decomposed into the projection integrals of the density distribution of iodine- and water-equivalent materials. After image reconstruction, the density images (central column in Figure 4) were combined together to generate the monochromatic image (right column in Figure 4). Note that the thorax phantom contains muscle, which is not pure water-equivalent, thus there are some residuals in the background of the iodine density image, although difficult to see in Figure 4 given the display range utilized.

A previous study indicated that there is an optimal energy level at which the monochromatic image has the highest SNR [4]. In this study we followed the results in literature and chose 70 keV as the energy level for the monochromatic image.

Figure 5 shows the monochromatic images of various sizes of the circular water phantom acquired with kVp switching (top row), ideal photon counting (middle row) and photon counting with N_0 =10 Mcps/mm^2 (bottom row). The tube current was 200 mA at 80 kVp and was scaled at 140 kVp for different phantom sizes to match the projection noise. Note that severe pile-up artifacts occur in the photon-counting CT image with N_0=10 Mcps/mm^2.

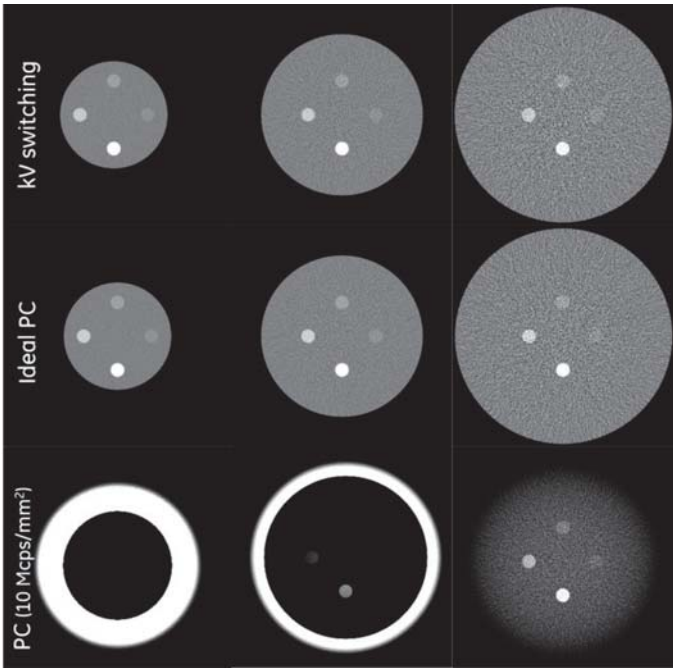


Figure 5 Monochromatic images of small (left column), medium (middle column) and large (right column) circular water phantoms using kVp switching (upper row), ideal photon counting (middle row) and photon counting with N_0=10 Mcps/mm^2 (bottom row).

Table 1 SNR in monochromatic images (Normalized to kVp switching)

	Circular		*Thorax*		*Abdomen*	
	kVp Switching	**Ideal PC**	**kVp Switching**	**Ideal PC**	**kVp Switching**	**Ideal PC**
Small	1.00	1.17	1.00	1.15	1.00	1.03
Medium	1.00	1.13	1.00	1.08	1.00	1.05
Large	1.00	1.08	-	-	-	-

Table 1 gives the quantitative results of SNR values in different phantom types and phantom sizes. All results were normalized to the SNR of kVp switching. The SNR for PC with pile-up effect was not included since the images were corrupted as shown in Figure 5. The results for large thorax and large abdomen phantoms were not included due to the streak artifacts caused by photon starvation.

B. Ideal PC vs. pile-up PC

As discussed in Section II.D we assume that pile-up in the air scan can be calibrated by scaling and averaging air scans at low mA. Four ROI‟s of iodine rods in each phantom were used to calculate the decomposition error according to the equations in Section II.E. The errors calculated for each ROI were then averaged for the results shown in Table 2.

Table 2 Decomposition error in medium sized thorax phantom

	Pile-up PC			Ideal PC
mA	**1 Mcps**	**10 Mcps**	**100 Mcps**	**∞ Mcps**
100	80.08%	2.09%	3.62%	1.49%
200	151.97%	8.98%	2.80%	1.82%
400	262.55%	29.01%	2.34%	1.57%
800	326.57%	64.28%	2.20%	1.37%

The decomposition errors from PCCT, including pile-up effects, for circular phantoms of different sizes are shown in Figure 6. For all phantoms, the errors in the monochromatic images from ideal PCCT were below 3% (resulting from model error in our decomposition methodology). When the detector"s maximum periodic rate N_0 is 100 Mcps, the pile-up effect did not have significant impact on decomposition accuracy (error < 4%) except for small circular phantom when tube current was above 400 mA. When N_0 is 10 Mcps, the decomposition error caused by the pile-up effect was significant for about half of the simulated cases. But for the large phantom using 200 mA or less, the results were still relatively accurate (error < 5%). Table 2 shows that at 100 mA the detector with N_0 = 10 Mcps outperformed the detector with N_0 = 100 Mcps in terms of decomposition error, which indicates that at low photon flux the pile-up for both cases were limited and the error was mainly due to the decomposition model. When N_0 is 1 Mcps, the decomposition error was above 50% for most of the cases.

IV. DISCUSSION

The evaluation scheme in this study is focused on dual-energy applications that are available today. It did not take into account the potential of some new technologies such as the recent development of novel nano-particle contrast agents. Specific imaging of such high-Z materials is possible if the energy threshold is chosen at the K-edge of that material, often referred to as „K-edge imaging" [10].

Unlike the conventional dual-energy technique, K-edge imaging often requires the collection of data in multiple energy bins (>2) and utilizes more advanced MD algorithms such as the Maximum Likelihood (ML) method. Previous studies indicated improved SNR in decomposed images when K-edge processing techniques are utilized in material decomposition [10]. However, it is still challenging to formulate nano-particles with high loadings of high-*Z* elements that are also nontoxic and able to be cleared from the body in a timely fashion [13].

V. SUMMARY

In summary, the impact of pile-up effect for photon-counting detectors was dependent on the detector"s maximum periodic rate (N_0). When N_0 was 1 Mcps/mm^2, the pile-up effect corrupted the images (with decomposition error greater than 50%) in most of the cases. For N_0 of 10 Mcps/mm^2, the decomposition error caused by the pile-up effect was

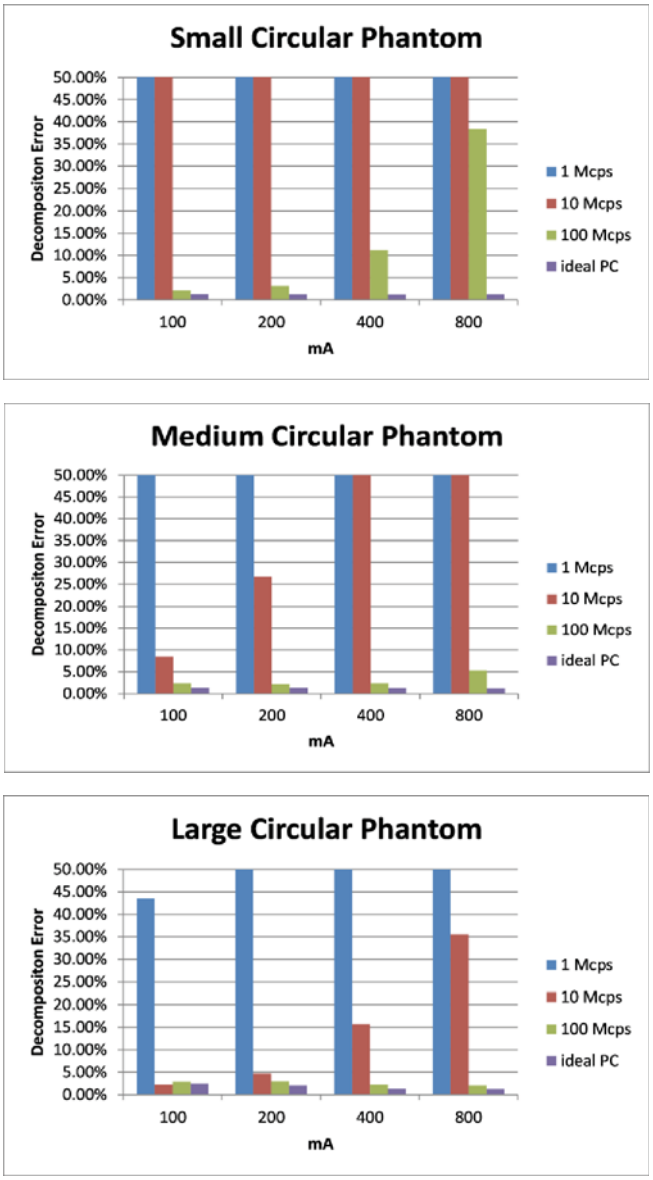


Figure 6 Decomposition error in small, medium and large circular phantoms for a tube current of100 / 200 / 400 / 800 mA. The detector"s maximum periodic rate $N_0$was 1/ 10 / 100 / ∞ Mcps / mm^2.

significant for large photon flux (large tube current, e.g. 400mA and 800mA, see table 2 and Fig. 6). For N_0>100 Mcps/mm^2, the pile-up effect did not have significant impact on decomposition accuracy (error < 4%) except for small circular phantom when the tube current was above 400 mA. Fast kVp switching methods perform favorably when compared to the current state of the art in photon-counting detectors; photon-counting CT systems using a reasonable value of N_0 (10 Mcps/mm^2) does not provide sufficient decomposition accuracy for clinical usage. The dual-energy performance of ideal photon-counting CT is about 5%-15% higher than that of the simulated kVp switching mode in terms of SNR in monochromatic images. However, development of photon-counting detectors with high counting rates, which are needed to achieve this potential, is a formidable challenge. The count-rate performance and clinical scan conditions are essential in evaluating the potential of a photon-counting system for routine clinical imaging.

REFERENCES

[1] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system," *British Journal of Radiology*, vol. 46, pp. 1016-1022, 1973.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Physics in Medicine and Biology*, vol. 21, pp. 733-744, 1976.

[3] F. Kelcz, P. M. Joseph, and S. K. Hilal, "Noise considerations in dual energy CT scanning," *Medical Physics*, vol. 6, pp. 418-425, 1979.

[4] W. A. Kalender, W. H. Perman, J. R. Vetter, and E. Klotz, "Evaluation of a prototype dual-energy computed tomographic apparatus. I. Phantom studies," *Medical Physics*, vol. 13, pp. 334-339, 1986.

[5] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Sü, M. Grasruck, K. Stierstorfer, B. Krauss, and R. Raupach, "First performance evaluation of a dual-source CT (DSCT) system," *European Radiology*, vol. 16, pp. 256-268, 2006.

[6] P. M. Shikhaliev, T. Xu, and S. Molloi, "Photon counting computed tomography: Concept and initial results," *Medical Physics*, vol. 32, pp. 427-436, 2005.

[7] P. M. Shikhaliev, "Tilted angle CZT detector for photon counting/energy weighting x-ray and CT imaging," *Physics in Medicine and Biology*, vol. 51, pp. 4267-4287, 2006.

[8] P. M. Shikhaliev, "Computed tomography with energy-resolved detection: a feasibility study," *Physics in Medicine and Biology*, vol. 53, pp. 1475-1495, 2008.

[9] P. M. Shikhaliev, "Energy-resolved computed tomography: first experimental results," *Physics in Medicine and Biology*, vol. 53, pp. 5595-5613, 2008.

[10] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multibin photon counting detectors," *Physics in Medicine and Biology*, vol. 52, pp. 4679-4696, 2007.

[11] S. Wang, D. Harrison, V. Lobastov, and J. E. Tkaczyk, "Pulse pileup statistics for energy discriminating photon counting x-ray detectors," *Medical Physics*, vol. 38, pp. 4265-4275, 2011

[12] De Man, Bruno, et al. "CATSIM: a new computer assisted tomography simulation environment." *Medical Imaging*. International Society for Optics and Photonics, 2007.

[13] Lin, Wenbin, et al. "Zr-and Hf-based nanoscale metal-organic frameworks as contrast agents for computed tomography." *Journal of Materials Chemistry,* vol. 22, pp. 18139–18144, 2012.

Practical Noise Correlation Modeling for Fluence Field Modulated Computed Tomography

Steven Bartolac and David Jaffray

***Abstract* ---- Fluence field modulated CT (FFMCT) proposes a novel approach for effective management of the tradeoff between radiation dose and image quality in CT by allowing the pattern of X-ray fluence incident on the patient to change as the source rotates about the detector. The specific modulation pattern can be optimized for the imaging task given some *a priori* information about the patient. The purpose of this work is to empirically model the noise response of the detector to allow for more accurate noise prediction when optimizing modulation patterns for prescribed image quality criteria. Local measurements of variance in an FFMCT reconstruction (generated using modulated projections synthesized from experimentally acquired projection sets) were compared to the predicted outcome and found to agree within 3%. This work highlights the importance of detector specific noise modeling for accurate noise prediction and image reconstruction in FFMCT. Significant dose reductions observed also support the advent of FFMCT in practice.**

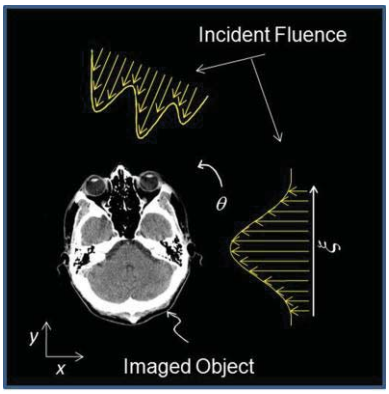


FIGURE 1 : Schematic of FFMCT illustrating that the incident pattern of fluence can change both across the field of view and between projections.

I. INTRODUCTION

A recent increase in concern over radiation due to CT procedures has largely emerged due to several publications over the last decade that have indicated both that the number of CT scans is rising substantially each year (on the order of 10% per annum[1]), and that the radiation dose to patients may have non-negligible lifetime attributable risk (LAR) of cancer[2, 3], especially when patients receive multiple CT scans. Appropriate selection of imaging parameters in CT inherently involves carefully balancing the tradeoff between the desired image quality (e.g. image noise) and the radiation delivered to the patient.

Limited modulation of the incident X-ray beam is already implemented in current CT units for decreasing dose to the patient while maintaining or bettering image quality compared to unmodulated approaches. The onset of automatic exposure control [4] combined with bow-tie filters [5-7], have made great strides towards more efficient management of the X-ray fluence. However, the ideal allocation of X-ray fluence would take into account both patient-specific anatomy as well as the imaging task, which suggests a filter that can be changed dynamically with respect to the X-ray tube source. Previous work [8, 9] has shown that optimizing the incident X-ray fluence, allowing its pattern to change spatially across the X-ray beam, and independently for each projection, may improve or maintain image quality where it is required, and allow for poorer image quality elsewhere in order to significantly decrease radiation dose to the patient.

An illustration of the proposed technique is shown in Figure 1. The methodology shares parallels with fluence planning and delivery used in intensity modulated radiation therapy (IMRT), and so was previously referred to as intensity modulated CT (IMCT). However, to avoid confusion with other tube current modulation techniques and to emphasize that changes can occur over the entire incident fluence field, the approach is now referred to as fluence field modulated CT (FFMCT).

In previous simulation studies, a parallel ray model and simplistic noise model were used to study the benefits of FFMCT to first order. The purpose of this work is to empirically model detector noise correlations for prediction of the noise in the reconstructed images, using real experimental data. This work is an extension of and builds in part on work presented at the Second CT Meeting [10].

Steven Bartolac is with the Department of Medical Biophysics, University of Toronto, 610 University Ave., Toronto, Ontario M5G 2M9 CANADA. (Tel: (416) 946-4501. Email: steve.bartolac@rmp.uhn.on.ca)

David Jaffray is with the Radiation Medicine Program, Princess Margaret Hospital/Ontario Cancer Insitute, 610 University Ave., Toronto, Ontario, M5G 2M9 CANADA. (Email: david.jaffray@rmp.uhn.on.ca)

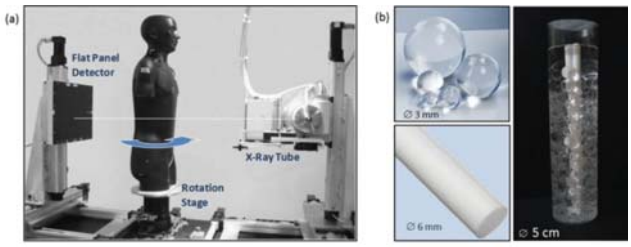


FIGURE 2: (a) Photo of experimental cone-beam CT unit. In this arrangement the x-ray tube and the detector are stationary, and the phantom rotates about a central axis on a rotation stage. (b) The phantom design used in this experiment consists of a 5 cm acrylic tube filled with acrylic spheres of varying sizes, polytetrafluoroethylene rods, and water.

II. METHODS

The proposed method for FFMCT requires first specifying target image quality objectives and/or dose constraints for the patient. The incident fluence pattern is then optimized as a function of angle and detector position using the preferred optimization algorithm. Implicit to this method is that some *a priori* information of the object is available; this information could be provided from a previous scan or could be inferred from a patient population model. Details regarding the methodology used for generating modulation profiles to meet prescribed image quality characteristics are found elsewhere [10, 11].

The present experiments were performed using an experimental cone-beam CT unit, consisting of an amorphous silicon flat panel detector (Paxscan 4030A, Varian, Palo Alto) with 194 μm pixel pitch, and a 600 kHU x-ray tube (Rad-94, Varian, Palo Alto). In this arrangement, the X-ray tube and detector are fixed in position, with the phantom mounted on a precision rotation stage, as shown in Figure 2.

The phantom imaged was a small cylindrical water phantom containing high contrast polytetrafluoroethylene rods and low contrast acrylic spheres of varying sizes (Figure 2). In order to generate a projection with a modulated fluence pattern, multiple scans of the object were taken of the phantom at different mAs settings (from 0.4 – 4 mAs per projection). A set of synthesized modulated projections were then created from the available projection sets, according to the modulation profile prescribed by the optimization method. The input model of the object was taken from a reconstruction of the object using a 1.25 mAs/projection setting.

The electronic noise in the detector is assumed to be fluence independent (i.e. the magnitude of electronic noise is the same irrespective of the mean photon counts at the detector). Electronic noise is measured by acquiring 100 *dark* field images (i.e. images with zero incident fluence) and measuring the variance in the pixels after subtracting the mean.

The variance in flat panel detectors is inherently less than what would be predicted purely from Poisson statistics. Siewerdsen et al.[12] describe in detail the various stages of detector readout and the associated gain and blurring that cause this effect. Though accurate modeling of the different factors involved is quite complex, a simplified empirical model describing the variance in a detector reading, based on the results of Siewerdsen et al.[12], was used instead:

$$\mathrm{var}(N) = aN + d^2 \qquad (1.1)$$

where the factor a is an empirically determined parameter meant to account for the reduced variance of the pixels due to blurring effects in the detector, and d is the measured standard deviation of the noise in the dark fields (i.e. "electronic noise").

A variance map based on the approach by Zhu et al. [13] was used in order to predict the noise distribution within the reconstructed image. This approach was derived for the fan-beam case with a flat panel detector, which corresponds with the present experiment. Further, it explicitly includes the transfer function for the noise which describes the nature of the noise correlation between pixels. From Zhu et al.'s algorithm, the variance in the image reconstruction data is approximated by

$$\mathrm{var}(\mathrm{f}(\vec{\mathrm{r}}) \simeq \sum \frac{1}{l^4(\vec{r})} B((\frac{w^2}{N^2} v * (g * h)^2 \Delta u, \theta)\Delta\theta^2 \qquad (1.2)$$

where B represents the filtered backprojection operator, l and w are weighting terms applied in fan-beam filtered backprojection, Δu is the pixel width at the detector, $\Delta\theta$ is the angular increment between projections, v is the variance in the projection data, and g is the transfer function governing the noise correlations between detector pixels. Dependencies on angular and detector position have been dropped to simplify the equation. Eq. (1.2) was derived assuming a white noise field, with variance v , that is subsequently convolved with the transfer function g. Therefore, v for the Poisson component would not be described by aN as defined in Eq. (1.1), since this value would represent the variance after convolution with g. Eq. (1.1) is therefore not convenient since we cannot directly measure v . However, once the function g is determined, the relationship between the variance of a hypothetical white noise field, and the variance of the same field after convolution with g can be found empirically in order to derive a proportionality constant. Therefore, for the Poisson component of noise

$$v = abN \qquad (1.3)$$

where b represents the empirically determined proportionality constant between noise fields before and after convolution

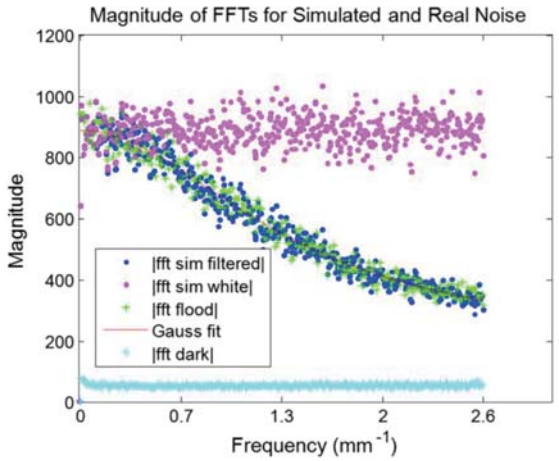


FIGURE 3: Average magnitudes of FFTs for simulated and real data. The simulated white noise signal shows as a flat spectrum in the FFT; after filtration, the simulated noise shows the same spectral characteristics as the real experimental data. The Gaussian fit shows good agreement with the shape of the FFT curve, and suggests a good choice for modeling the noise transfer function. A characteristic white spectrum is seen in the FFT of the dark fields (electronic noise). Note also that the magnitude of the electronic noise is also much less than the Poisson component.

with *g*. Since the electronic noise is independent of the Poisson noise, it can be considered separately. The proportionality constant and a separate transfer function can also be determined for the electronic noise component.

We note that in their paper, Zhu et al. use the transfer function *g* synonymously with the modulation transfer function (MTF). However, the transfer function for noise modeled here is proportional to the root of the magnitude of the noise power spectrum (NPS). The MTF and NPS are related functions but need not share the same profile or shape; therefore in the present study *g* is not assumed to be the MTF as executed in the paper by Zhu et. al. Further, as noted above, the electronic noise and the Poisson noise may not exhibit the same correlations between pixels, and so *g* should be evaluated separately for each.

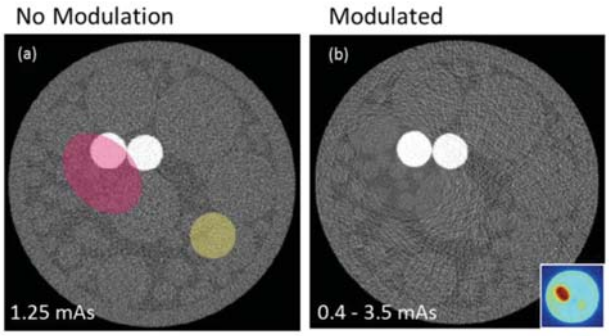


FIGURE 4: (a) Reconstruction of data acquired using a 1.25 mAs tube current setting. (b) Reconstruction of data using synthesized modulated projections, where the tube current setting ranged from 0.4 – 4 mAs. The bottom right corner in the modulated projection data set illustrates the predicted SNR distribution. Colour washed regions in (a) correspond to regions of high and intermediate image quality predicted in the FFMCT case for visual comparison.

Analysis of the electronic noise was performed by acquiring 100 dark acquisitions without any fluence incident on the panel. Analysis of correlations is performed by calculating the discrete linear Fourier transform of each dark field using the Matlab fast Fourier transform (FFT) function, and observing the average magnitude of the Fourier components. As shown in the results section, the electronic noise was found to behave as a white noise field, so no further steps in analysis were required in terms of calculating the transfer function for electronic noise.

These steps were repeated for *flood* fields (no object in the field of view), with tube current settings ranging from 0.4 – 1 mAs. The flood fields contain both Poisson and electronic noise. Variances of independent noise sources are added in quadrature, so the electronic noise component can be separated from the Poisson noise component by first squaring the mean absolute FFT of the flood fields and then subtracting the square of the FFT of the electronic noise. Taking the square root of the result provides a measure of the FFT of the noise due to Poisson statistics alone. Finally, the transfer function was modeled in the Fourier domain by fitting a Gaussian curve to the FFT. In order to calculate the proportionality constant in Eq. (1.3) white noise fields were generated and then convolved with *g*. The variance was calculated before and after the convolution and the results compared to derive *b*. Calculation of *g* in the real domain would require an inverse FFT of the fitted function in the Fourier domain. However, practically, the convolution in Eq. (1.2) of *g* and *h* is done directly in the Fourier domain and therefore this last step is not required.

A variance map for a high resolution FFMCT image was generated using Eq. (1.2). Variance measurements in the FFMCT reconstructed data were then compared to the predicted variance map for evaluation of the accuracy of the fan-beam prediction model in conjunction with the noise modeling of the detector specific to the imaging system used.

III. RESULTS AND DISCUSSION

Figure 3 shows the average magnitude of the FFT for a measured flood field (1 mAs) fitted with a Gaussian curve. As shown, the fit agrees well with the measured data and is used to model the magnitude of the FFT of the transfer function *g* required in Eq. (1.2). Figure 3 also shows the amplitude of the FFT of a simulated white noise field before and after convolution with *g*. Agreement of the simulated and real noise distributions gives confidence that the modeled transfer function accurately characterizes the correlations in the noise. The form of the normalized transfer function was also found to be consistent for different flood fields suggesting that the relative shape of *g* is not dependent on the magnitude of the incident fluence. Comparison of the variance of the white noise field prior to and after

convolution with g yielded a value of 2.12 for the proportionality constant b. The same result for b was found regardless of the magnitude of the white noise distribution. Also shown in Figure 3 is the average magnitude of the FFT of electronic noise (from acquired dark field data). Note that the FFT for the electronic noise is approximately flat across frequency bins. The transfer function was therefore assumed to be unity for the electronic noise, and the proportionality constant is also assumed to be 1 in that case.

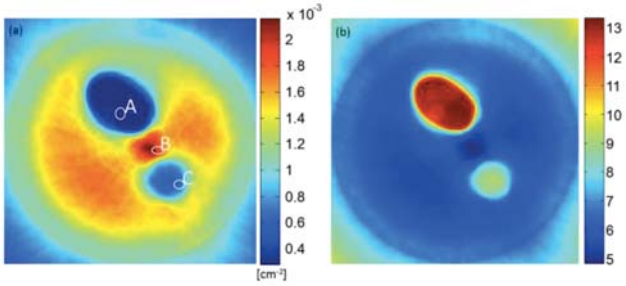


FIGURE 5: (a) Variance distribution predicted using the Eq. derived by Zhu et. al. The mean variance in regions A, B and C were compared to the actual measured variance in the experimental data and found to agree well within 5%. (b) Predicted ratio of water signal to noise.

Figure 4 shows a comparison of the reconstructions of the cylinder using synthesized FFMCT projections versus a reconstruction using an unmodulated beam and using a 1.25 mAs tube current setting. The FFMCT reconstruction shows better image quality within the prioritized region of interest for the FFMCT case, with reduced quality elsewhere, following closely with the predicted image quality map. Dose reduction in the FFMCT case was also found to be on the order of 40%.

A variance map for the high resolution data is shown in Figure 5 (a), with the corresponding water signal-to-noise ratio (SNR) distribution also shown in Figure 5 (b). Measurements of the variance within the regions indicated in Figure 5 (a) were made using the reconstructed data shown in Figure 4 (b). Results indicated agreement within 3% between the predicted and experimental results.

IV. CONCLUSION

The outcomes of this study further support that FFMCT could potentially be applied with success in real clinical CT systems, provided that a suitable method for modulation be found. Results show very good agreement between predicted and measured noise outcomes, as well as significant reductions in dose.

REFERENCES

[1] D. Sutton, "Population and patient risk from CT scans.," *Controversies and Consensus in Imaging and Intervention,* vol. 6, pp. 4-9, 2008.

[2] D. J. Brenner and E. J. Hall, "Computed tomography--an increasing source of radiation exposure," *N Engl J Med,* vol. 357, pp. 2277-84, Nov 29 2007.

[3] R. Smith-Bindman, "Is computed tomography safe?," *N Engl J Med,* vol. 363, pp. 1-4, Jul 1 2010.

[4] M. Kachelriess, *et al.*, "Strategies for Dose Reduction and Improved Image Quality in MSCT," in *Multidetector-Row CT of the Thorax*, U. Joseph Schoepf, Ed., Berlin: Springer, 2006, pp. 35-45.

[5] S. A. Graham, *et al.*, "Compensators for dose and scatter management in cone-beam computed tomography," *Med Phys,* vol. 34, pp. 2691-703, Jul 2007.

[6] N. Mail, *et al.*, "The influence of bowtie filtration on cone-beam CT image quality," *Med Phys,* vol. 36, pp. 22-32, Jan 2009.

[7] M. A. Hassan, *et al.*, "A radiotelemetry pill for the measurement of ionising radiation using a mercuric iodide detector," *Phys Med Biol,* vol. 23, pp. 302-8, Mar 1978.

[8] J. Smyth, *et al.*, "Evaluation of the quality of CT-like images obtained using a commercial flat panel detector system," *Biomed Imaging Interv J,* vol. 2, p. e48, Oct 2006.

[9] S. A. Graham, "Intensity Modulated Cone-Beam CT," M.Sc., Medical Biophysics, University of Toronto, Toronto, 2006.

[10] S. Bartolac and D. Jaffray, "Fluence field modulated computed tomography," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, Salt Lake City, 2012.

[11] S. Bartolac, *et al.*, "Fluence field optimization for noise and dose objectives in CT," *Med Phys,* vol. 38 Suppl 1, p. S2, May 2011.

[12] J. H. Siewerdsen, *et al.*, "A framework for noise-power spectrum analysis of multidimensional images," *Med Phys,* vol. 29, pp. 2655-71, Nov 2002.

[13] L. Zhu and J. StarLack, "A practical reconstruction algorithm for CT noise variance maps using FBP reconstruction," pp. 651023-651023, 2007.

A Non-Uniform Super-Resolution Compressive Sampling Method For Tomographic Imaging

Synho Do PhD[1], *W. Clem Karl PhD*[2], *Thomas Brady MD*[1], *Georges El Fahkri PhD*[1], *and Rajiv Gupta MD PhD*[1]

ABSTRACT

We present a new non-uniform sampling method for tomographic sensing systems that allows super-resolution image reconstruction. In conventional tomographic systems, the resolution is determined by the detector size and the angular sampling interval. In this work, we propose a non-uniform sampling scheme based on an Archimedean spiral sampling pattern that allows for an enhanced reconstruction resolution much greater than the detector size to be achieved. We present an associated model-based iterative image reconstruction approach that achieves this enhanced resolution. Initial simulation results demonstrated a many-fold resolution improvement in parallel beam geometry. In addition, we show that the proposed method can be implemented in helical cone-beam Multi-row Detector CT (MDCT) configurations by modifying the flying focal spot motion instead of the detector motion, allowing its application to current commercial systems.

Index Terms— Radon space, Sinogram, Super-resolution, Computed tomography, Iterative reconstruction, MDCT

1. INTRODUCTION

There have been many advances in the physics of medical tomographic imaging systems, especially in the advent of multi-detector row helical systems [1] [2] [3]. The resolution of the conventional tomographic systems is driven by the detector size and angular sampling. Historically these quantities have been constrained to a regular grid. Quarter detector off-set [4] and the use of a Flying Focal Spot (FFS) [5] [6] can improve the sampling density, and thus the resolution of CT systems.

Non-uniform sampling schemes based on general *k*-space trajectory studies [7] [8] [9] and encoding methods [10] [11] have been proposed in the domain of Magnetic Resonance Imaging (MRI) to improve speed [12] [13] and sampling density [14] [15]. These methods aim to reduce data sampling, and mitigate under-sampling artifacts and motion artifacts by combining the non-uniform sampling with advanced reconstruction methods.

In this paper, we propose a novel spiral sampling pattern for tomographic imaging problems based on fractional shifts of the system detectors on an Archimedean spiral on Radon Space (ASRS). We couple this non-uniform sampling pattern with a model-based reconstruction approach using an accurate system model [16] demonstrating the super-resolution properties of the resulting system. In addition, we show that the proposed method can be implemented in helical cone-beam Multi-row Detector CT (MDCT) configurations by modifying the flying focal spot motion instead of the detector motion, allowing its application to current commercial systems..

2. METHOD

The Archimedean or arithmetic spiral (AS) is the locus of points corresponding to the locations over time of a point moving away from a fixed origin with a constant speed along a line, which rotates with constant angular velocity [17]. The radial distance, r, follows the following function:

$$r = a + b\theta^{\frac{1}{k}} \tag{1}$$

where θ is the polar angle and a and b are constants. When k changes from -2 to 2, the spiral wraps tightly. Figure 1 shows the trajectory of (θ, r) in a Cartesian grid with $a = 0$, $b = 1$, and $0 \le \theta \le 540$.

We propose applying this trajectory as a sampling pattern for tomographic systems by shifting the detector system 1/n of the detector size for each of n angular sampling positions to create an Archimedean spiral (k = 1) sampling pattern on Radon space. The schematic diagram of such a fractional shift is illustrated in Figure 2-(a). At each angular sample, the detector system performs a fractional shift. This shift can be defined by the polar angle θ.

[1] Massachusetts General Hospital and Harvard Medical School, Boston, MA. E-mail: sdo@nmr.mgh.harvard.edu
[2] Boston University, Boston, MA

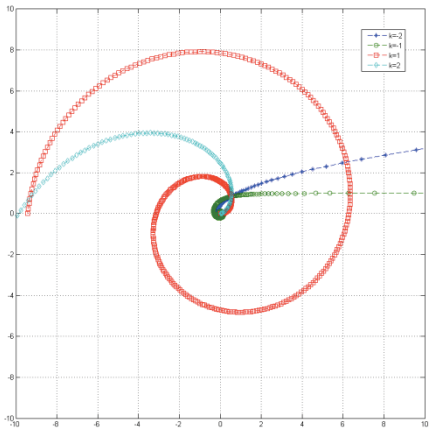


Figure 1. A generalized Archimedean spiral with k = [-2,-1,1,2]: Cartesian grid drawing of Archimedean spiral with a=0, b=1, $0 \le \theta \le 540$

The resulting sampling trajectory can be interpreted in the Sinogram domain. In Figure 2-(b), the detector sampling trajectories without shift are shown as straight solid lines while the sampling trajectories with shift are shown as broken lines. The broken lines are slanted in the Sinogram space which makes it possible to acquire sub-detector pixel information at each view. These tilted lines on the Sinogram are equivalent to ASs on Radon space.

This proposed method to produce an irregular sampling pattern on Radon space for each detector element is exactly described by the modified AS:

$$r = a + (b'/2n)\theta^{\frac{1}{k}} \qquad (2)$$

where a is the initial detector location, b' is the normalized detector span pitch, and n is rotation/180 degree. We used k=1 for our simulation. The modified AS can easily model multiple spirals from multiple detector elements.

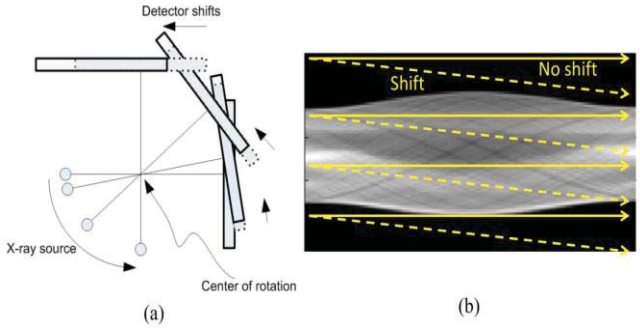


Figure 2. A diagram of fractional shift of detector systems. (a) Detector system shifts 1/n of detector size for n angular sampling position (b) In the Sinogram domain, we can compare sampling trajectories with detector shifts (broken line) and without shifts (solid line).

We define a geometric system model corresponding to the sampling pattern defined c.f. Eq. (2). By collecting variables in a vector-matrix equation we can write the set of observations as:

$$y = Tf \qquad (3)$$

where y is the set of projection observations, f is the set of unknown pixel values, and T is the matrix that relates the two to each other.

We produce an image as the solution of the following optimization problem:

$$\hat{f} = \arg\min \|y - Tf\|^2 + \lambda \|Df\|^p \qquad (4)$$

where D is a gradient operator and p=1. The details of implementation can be found in [18].

3. RESULTS AND DISCUSSION

To illustrate the proposed method, we initially simulate tomography for a 12x12 pixel field. The corresponding sinogram is created using only 2 detector elements covering the field of view and 100 angular sampling positions as shown in Figure 3-(a). This sinogram is generated by the proposed detector shifting scheme as shown in Figure 2.

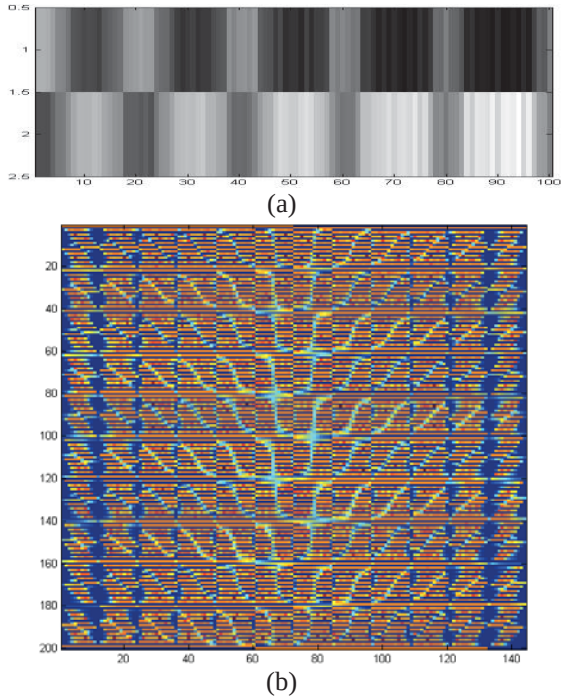


Figure 3, Sinogram and system matrix generated by proposed method: (a), Sinogram generated by 2 detector elements with detector shifts. (b) System matrix includes geometry of scanner and detector motion

The system matrix of this projection scheme (including detector shifting) is configured as a sparse matrix as shown in Figure 3-(b). The size of system matrix is 144 (12 x 12) by 200 (2 x 100). The system matrix is used for iterative image reconstruction with L-1 norm [18]. Figure 4 compares reconstructed images: (a) Ideal image (12 by 12), (b) Reconstructed image by conventional method, (c) Error in conventional image (true-conventional). (d) reconstructed image with the proposed method, and (e) error of new method (true-new). Note that the scale of difference map is very small in Figure 4-(e).

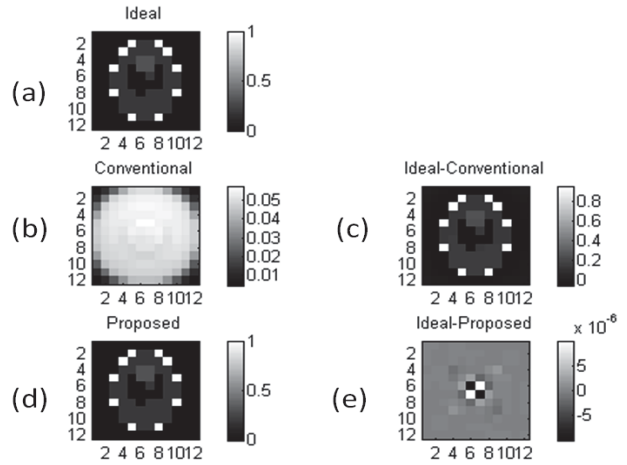


Figure 4, Image comparison: (a) Ideal image (12 by 12), (b) Reconstructed image by convention method, (c) Difference map between (a) and (b). (d) Reconstructed image with the proposed method, and (e) Difference map between (a) and (d).

Our shifting scheme produces a 6-fold resolution improvement over the native detector resolution. The sampling patterns on Radon space (i.e., a polar view of Radon space compared to Sinogram) are compared in Figure 5. The conventional sampling patterns with 12 and 64 detector elements are shown in Figure 5-(a) and (d). The reduced detector sampling patterns with a conventional method show coarse sampling trajectories in Figure 5-(b) and (e). In contrast, the proposed method can generated sparse, irregular patterns on Radon space as shown in Figure 5-(c) and (f). Both sparseness and irregularity are important components in compressed sensing theory [19, 20].

We also implemented the proposed ASRS method in helical cone beam clinical scanner geometry (Brilliance 64-MDCT, Philips Healthcare) with patient data. Instead of detector system motion, we simulated slowly changing the Flying Focal Spot location (small-step FFS). While the total distance traveled by the focal spot in small-step FFS is the same as it does in conventional FFS, it travels the distance in many smaller steps, rather than simply moving from end to end as it does in conventional FFS. However, the degree of total deviation is very close to the real FFS. This is a source motion encoding method compared to the detector motion encoding approach we described in the previous simulation. Both approaches generate similar spiral sampling patterns on Radon space even though it's more difficult to visualize spiral patterns in a 3D helical case.

4. CONCLUSION

In this paper, we present a novel Archimedean spiral sampling method on Radon space for a super-resolution image reconstruction. We implemented the exact form of modified Archimedean spiral equation in 2D and 3D cases. We proposed two approaches to generate Archimedean spiral patterns on Radon space: 1) fractional shifts of the detector and/or 2) X-ray focal spot motion encoding. In the noiseless case, we could easily achieve many-fold resolution improvement. In summary, the resolution of tomographic system is not limited by the detector size but by the sparseness of sampling on Radon space.

5. ACKNOLEDGEMENT

The authors gratefully acknowledge Kevin Brown, Philips Healthcare, for providing system information and constructive discussion.

6. REFERENCES

[1] M. Prokop, "General principles of MDCT," *European Journal of Radiology,* vol. 45, pp. S4-S10, 2003.

[2] T. Flohr, K. Stierstorfer, S. Ulzheimer, H. Bruder, A. Primak, and C. McCollough, "Image reconstruction and image quality evaluation for a 64-slice CT scanner with z-flying focal spot," *Medical physics,* vol. 32, p. 2536, 2005.

[3] T. Giesler, U. Baum, D. Ropers, S. Ulzheimer, E. Wenkel, M. Mennicke, W. Bautz, W. A. Kalender, W. G. Daniel, and S. Achenbach, "Noninvasive visualization of coronary arteries using contrast-enhanced multidetector CT: influence of heart rate on image quality and stenosis detection," *American Journal of Roentgenology,* vol. 179, pp. 911-916, 2002.

[4] P. J. La Rivière and X. Pan, "Sampling and aliasing consequences of quarter-detector offset use in helical CT," *Medical Imaging, IEEE Transactions on,* vol. 23, pp. 738-749, 2004.

[5] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Süβ, M. Grasruck, K. Stierstorfer, B. Krauss, and R. Raupach, "First performance evaluation of a dual-source CT (DSCT) system," *European radiology,* vol. 16, pp. 256-268, 2006.

[6] M. Kachelrieß, M. Knaup, C. Penßel, and W. A. Kalender, "Flying focal spot (FFS) in cone-beam CT," *Nuclear Science, IEEE Transactions on,* vol. 53, pp. 1238-1247, 2006.

[7] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magnetic Resonance in Medicine,* vol. 46, pp. 638-651, 2001.

[8] E. Adalsteinsson, P. Irarrazabal, S. Topp, C. Meyer, A. Macovski, and D. M. Spielman, "Volumetric spectroscopic imaging with spiral-based k-space trajectories," *Magnetic Resonance in Medicine,* vol. 39, pp. 889-898, 2005.

[9] G. H. Glover, "Simple analytic spiral K-space algorithm," *Magnetic Resonance in Medicine,* vol. 42, pp. 412-415, 1999.

[10] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine,* vol. 42, pp. 952-962, 1999.

[11] K. P. Pruessmann, "Encoding and reconstruction in parallel MRI," *NMR in Biomedicine,* vol. 19, pp. 288-299, 2006.

[12] C. H. Meyer, B. S. Hu, D. G. Nishimura, and A. Macovski, "Fast spiral coronary artery imaging," *Magnetic Resonance in Medicine,* vol. 28, pp. 202-213, 2005.

[13] C. Ahn, J. Kim, and Z. Cho, "High-speed spiral-scan echo planar NMR imaging-I," *Medical Imaging, IEEE Transactions on,* vol. 5, pp. 2-7, 1986.

[14] R. D. Hoge, R. K. S. Kwan, and G. Bruce Pike, "Density compensation functions for spiral MRI," *Magnetic Resonance in Medicine,* vol. 38, pp. 117-128, 2005.

[15] J. R. Liao, J. M. Pauly, T. J. Brosnan, and N. J. Pelc, "Reduction of motion artifacts in cine MRI using variable-density spiral trajectories," *Magnetic Resonance in Medicine,* vol. 37, pp. 569-575, 2005.

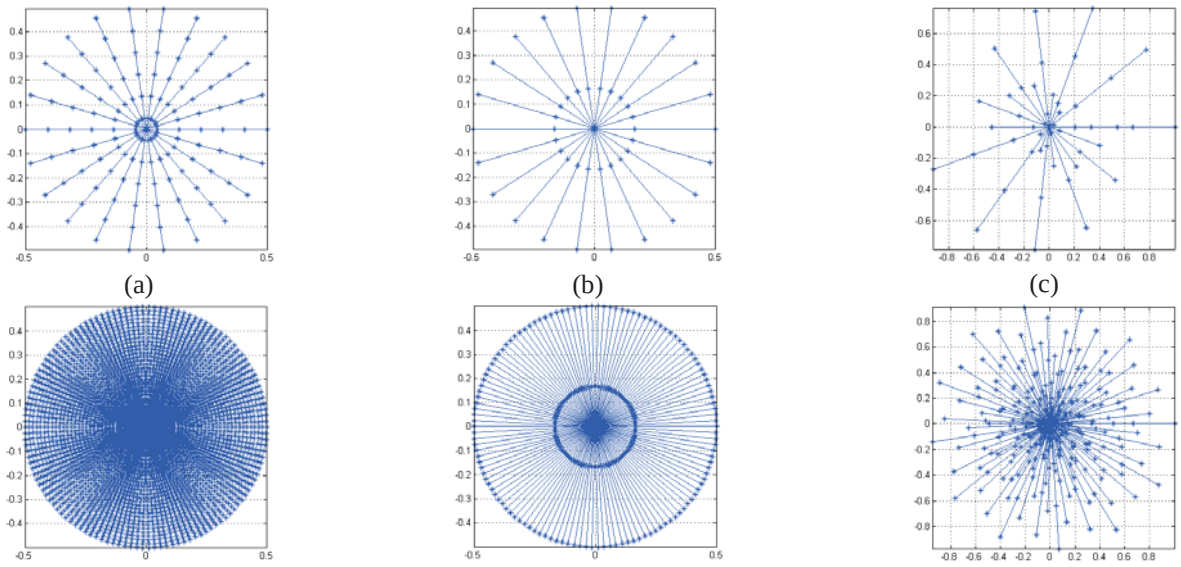
[16] S. Do, S. Cho, W. C. Karl, M. K. Kalra, T. J. Brady, and H. Pien, "Accurate model-based high resolution cardiac image reconstruction in dual source CT," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, 2009, pp. 330-333.

[17] E. H. Lockwood, *A book of curves*: Cambridge University Press, 1961.

[18] S. Do, W. C. Karl, Z. Liang, M. Kalra, T. J. Brady, and H. H. Pien, "A decomposition-based CT reconstruction formulation for reducing blooming artifacts," *Physics in Medicine and Biology,* vol. 56, p. 7109, 2011.

[19] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on,* vol. 52, pp. 1289-1306, 2006.

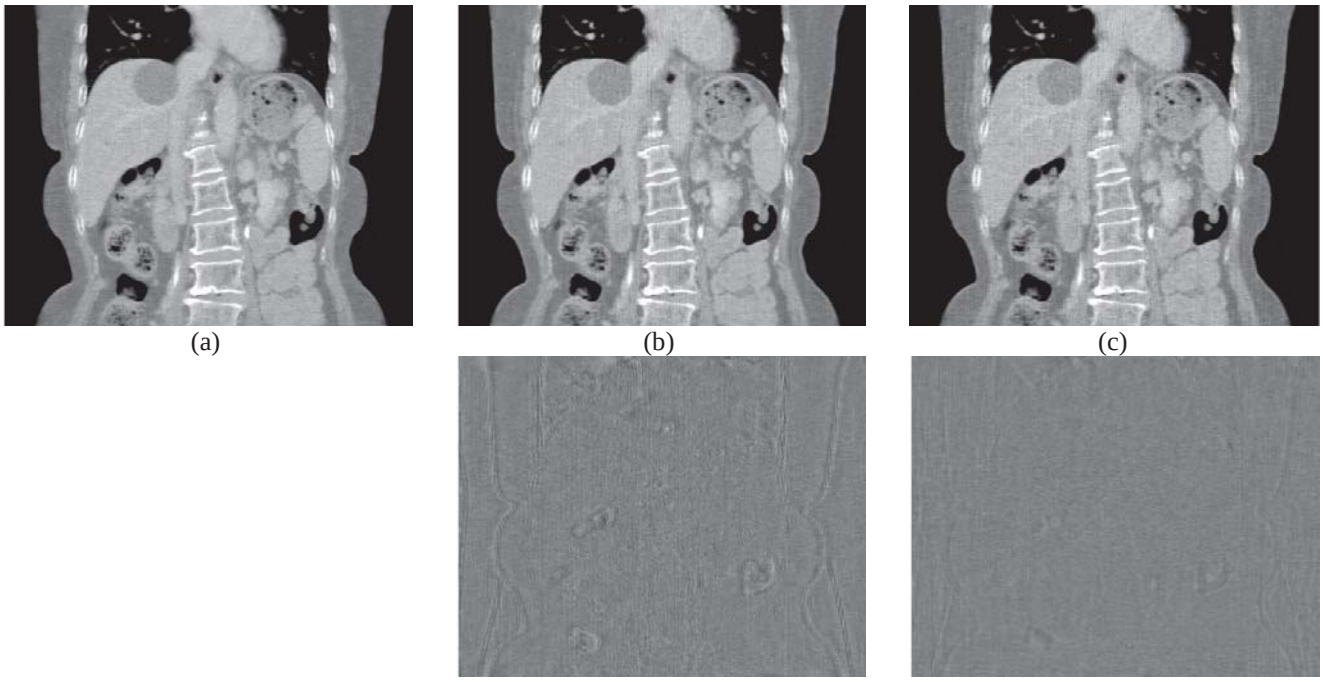
[20] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique,* vol. 346, pp. 589-592, 2008.



(a) (b) (c)

(d) (e) (f)

Figure 5, Sampling pattern comparison: (a) conventional sampling pattern with 12 detector elements and 12 angular positions, (b) conventional sampling pattern with 4 detector elements and 12 angular positions, (c) the proposed method with 4 detector elements and 12 angular positions. (d) convention sampling pattern with 64 detector elements and 64 angular position (e) conventional sampling pattern with 4 detector and 64 angular positions, and (f) the proposed method with 4 detector elements and 64 angular positions.



(a) (b) (c)

(d) (e)

Figure 6, Reconstructed image comparison (Coronal Views): (a) Original image (dx=dy=0.75, dz=0.625) displayed in contrast window [-500 500] HU. The image is reconstructed by 672-column and 64-row detector system with 18444 angular views (b) images reconstructed by the proposed method with three times bigger detector elements covering the same field of view so that we used 226-column and 64-row detector system with 18444 angular views, (c) The images are reconstructed with three times bigger detector elements with 50% of angular views so that we used 226-column and 64-row detector system with 9222 angular views, (d) shows the difference map between (a) and (b), and (e) shows the difference map between (b) and (c)

Comparison of the Scanning Linear Estimator and 3D ROI Operator for Quantitative ^{111}In-Octreotide SPECT Imaging

Arda Könik, Meredith Kupinski, P. Hendrik Pretorius, Michael A. King, and Harrison H. Barrett

***Abstract*— In quantitative PET and SPECT imaging signal activity is typically estimated from calculations in a 3D region of interest (ROI) of the reconstructed slices. However, unpredictable bias arising from the null functions of the imaging system affects ROI estimates. It has been shown that the scanning linear estimator (SLE), which operates on the raw projection data, is an unbiased estimator of activity when the size and location of the signal are known. In this work, we compared the ROI and SLE methods through analytic simulations of a realistic NCAT human phantom. The task was to estimate the signal activity of a lesion in an ^{111}In-Octreotide SPECT study when the size and location of the signal are known. Quantification in terms of the total count ratio (TCR) was determined for signals of varying sizes and contrast levels. The performance of SLE and ROI were compared for 3 conditions: 1) noiseless measurements and a constant background, 2) noisy measurements and a constant background, and 3) noisy measurements and a variable lumpy background. In summary, we showed that the SLE outperformed the ROI method by always achieving a lower bias with a lower variance except for the smallest lesion.**

***Index Terms*— Quantitative SPECT, ROI, SLE**

I. Introduction

In quantitative emission tomography, the activity in a lesion is typically estimated based on the counts in a 3D region of interest (ROI) of the reconstruction. However, ROI methods are subject to unpredictable bias arising from the null functions of the imaging system and the object. It has been shown that scanning linear estimation (SLE), which is calculated from the raw projection data, provides unbiased results (i.e., the average estimate equals the true value) for activity estimation when the signal size and location are known [1, 2].

To demonstrate a clinical application of SLE we have considered ^{111}In-Octreotide SPECT studies. We compared the performance of the ROI and SLE methods through an analytical imaging simulation of a realistic NCAT phantom [3]. The task was to estimate the activity of a spherical lesion of known diameter and location that has been added to the non-targeted activity in the liver. Estimation performance was compared for varying signal sizes and contrast levels. Lesion sizes reported in the literature varied from ~1-6 cm (mostly 1-2.5 cm range) [4] with uptake ratio (tumor / liver background) of ~ 1-7 [5]. In this work, 4 signal diameters from 1-3 cm were simulated and for each size the contrast ratio varied from 1-10.

II. Methods

We compared the performance of ROI and SLE for a range of signal sizes and activity levels in three different conditions: 1) noiseless measurements and a constant background, 2) noisy measurements and a constant background, and 3) noisy measurements and a variable background. For case 2 and 3, statistics are reported from 100 independent sample images.

In the following sections, preparation of the phantom, the imaging simulation, and SLE and ROI estimation methods are described.

A. Phantom and Projection Data

We generated NCAT activity and attenuation phantoms (matrix size: 256x256x256 and pixel size: 0.2332 cm) with slices outside the liver region set to zero. The coefficients of the attenuation phantom were based on the 171 keV emission photons of ^{111}In radionuclide (245 keV emission was not considered). In the activity phantom simulated signals were added to the liver. Varying the signal diameter from: 1, 1.5, 2 and 3 cm and the contrast ratios from 1-10, we created an ensemble of signals, where the contrast ratio (CR) is defined as CR = (Signal – Background) / Background.

Using an analytical projector with photon attenuation (without scatter effect) and distant-dependent collimator blurring (medium energy collimator), projections were obtained over 120 angles covering 360 degrees around the phantom. Then, 256x256x120 projection data were resized to 128x128x120 and scaled to ~ 5 million total counts, which represented the average count level of an ^{111}In-Octreotide SPECT study.

Manuscript received April 15, 2013. This work was supported by NIH grants P41 EB002035-13, Center for Gamma-Ray Imaging, and R01 EB000803-21, SPECT Imaging and Parallel Computing. The contents are solely the responsibility of the authors and do not represent the official views of the NIH.

Arda Könik is with the Radiology Dept, UMass Medical School, Worcester, MA 01655, (e-mail: arda.konik@umassmed.edu).

Meredith Kupinski is with the College of Optical Sciences and Center for Gamma-Ray Imaging, Dept. of Medical Imaging, University of Arizona, Tucson, AZ 85724, (e-mail: meredith@optics.arizona.edu).

P. Hendrik Pretorius is with the Radiology Dept, UMass Medical School, Worcester, MA 01655, (e-mail: hendrik.pretorius@umassmed.edu).

Michael A. King is with the Radiology Dept, UMass Medical School, Worcester, MA 01655, (e-mail: michael.king@umassmed.edu).

Harrison H. Barrett is with the College of Optical Sciences and Center for Gamma-Ray Imaging, Dept. of Medical Imaging, University of Arizona, Tucson, AZ 85724, (e-mail: barrett@radiology.arizona.edu).

B. Scanning Linear Estimator (SLE) Method

The SLE of activity was calculated from the raw projection data using the following estimation rule, which was derived in [1, 2].

$$\hat{\alpha}_{SLE}(g) = \frac{s^t K^{-1}_{g|\bar{\alpha}} (g - \bar{b})}{s^t K^{-1}_{g|\bar{\alpha}} s} \quad (1)$$

s : noise - free projection data of unit activity signal and s^t : $transpose(s)$

$K_{g|\bar{\alpha}} = \bar{K}_n + \bar{K}^{bkgd}_{\bar{g}}$ (ensemble average of covariance matrix)

$K^{-1}_{g|\bar{\alpha}}$: inverse ($K_{g|\bar{\alpha}}$)

g : noisy projection data of the background and the signal

$\bar{b}$: noise - free projection data of the background

This activity estimate was compared to the truth using the total count ratio, TCR_{SLE} % = 100 x SLE / TRUE.

The average value of the SLE is equal to the true value of activity

$$\left\langle \hat{\alpha}_{SLE}(g) \right\rangle_{g|\alpha} = \frac{s^t K^{-1}_{g|\bar{\alpha}} (\alpha s + \bar{b} - \bar{b})}{s^t K^{-1}_{g|\bar{\alpha}} s} = \alpha \quad (2)$$

where α is the true signal activity and $\bar{g} = \alpha s + \bar{b}$ is the projection data averaged over noise and background realizations. When the uptake in the liver is random, the covariance matrix becomes non - diagonal and poses a challenging computational task to SLE implementation.

C. Region of Interest (ROI) Method

For the ROI studies, the projection data were reconstructed using an iterative reconstruction algorithm (OSEM: 5 iterations and 30 subsets) with attenuation correction, resolution compensation and without post-filtering. With the known signal size and location, the total count ratio (TCR) [6] of the reconstructed and true object was obtained within the 3D ROI (signal support).

$$TCR_{ROI}\% = 100 \cdot \frac{\sum_{sphere} (G_{recon} - \bar{B}_{recon})}{\sum_{sphere} (G_{true} - B_{true})} \quad (3)$$

G_{recon} : Noisy reconstruction of the background and signal

$\bar{B}_{recon}$: Noise - free reconstruction of the background

G_{true} : True object with signal

B_{true} : True background object (without signal)

where $G_{recon} \to \bar{G}_{recon}$ for the noiseless case.

III. Results

Fig. 1 shows sample images for the constant (size: 1-3 cm) and lumpy background (size: 2 cm) studies at CR=5. At this contrast level, 1 and 1.5 cm signals are difficult to identify visually even in the noiseless projections. While TCR_{SLE} was 100% for all noiseless cases, TCR_{ROI} ranged from 48-59% for 1 cm and 65-85% for 1.5 cm diameter signals; respectively. At 2 and 3 cm, ROI estimations were closer to the true values and with less variation, especially among higher contrast levels. The results for the noiseless ROI measurements are summarized in Fig. 2.

SLE and ROI mean and standard deviation (SD) of TCR, based on 100 independent samples, are presented in Fig. 3 and Fig. 4; respectively. SLE was unbiased (i.e., mean TCR ~100%) in all cases except for the lowest contrast ratio for the diameter of 1 and 1.5 cm. This is due to the greater variance in SLE estimates of small and weak signals. The ROI estimation bias depended strongly upon the signal size and activity level. Mean TCR_{ROI} values ranged from 50-100% over the ensemble of signals. As the signal size increased, the performance of both estimation methods improved. When the signal size was 3 cm, the average value of TCR_{ROI} was near the 100% level.

IV. Discussion

The results showed that the SLE performed better than the ROI method by having a much lower bias and a somewhat lower variance across all sampled lesion parameters except for the smallest signal size of 1 cm diameter which is not typical for liver lesions [4, 5]. The bias of ROI estimation is highly dependent upon both signal contrast and size, which has indications for its utility in monitoring response to treatment.

V. Conclusion

We showed with a series of ^{111}In-Octreotide SPECT simulations SLE outperformed the ROI method with unbiased estimations and lower variability in the results. In future work, we will be investigating the performance of the SLE method in various liver textures, more realistic simulations, and usage of patient data to estimate the image covariance matrix.

References

[1] M. K. Whitaker, E. Clarkson, and H. H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods," *Opt Express*, vol. 16, pp. 8150-73, 2008.

[2] M. K. Whitaker, E. Clarkson, and H. H. Barrett, "Bias in ROI estimators and an unbiased solution," in *NSS Conf. Record*, 2008, pp. 5332- 5334

[3] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Med Phys*, vol. 37, pp. 4902-15, 2010.

[4] O. Schillaci, et al "Somatostatin receptor scintigraphy in liver metastasis detection from gastroenteropancreatic neuroendocrine tumors," *J Nucl Med*, vol. 44, pp. 359-68, 2003.

[5] D. K. Kontogeorgakos, et al "Patient-specific dosimetry calculations using mathematic models of different anatomic sizes during therapy with In-111-DTPA-D-Phe(1)-octreotide infusions after catheterization of the hepatic artery,"*JNM* , vol. 47, pp.1476-1482, 2006.

[6] P. H. Pretorius, M. A. King, T. S. Pan, D. J. de Vries, S. J. Glick, and C. L. Byrne, "Reducing the influence of the partial volume effect on SPECT activity quantitation with 3D modelling of spatial resolution in iterative reconstruction," *Phys Med Biol*, vol. 43, pp. 407-20, Feb 1998.

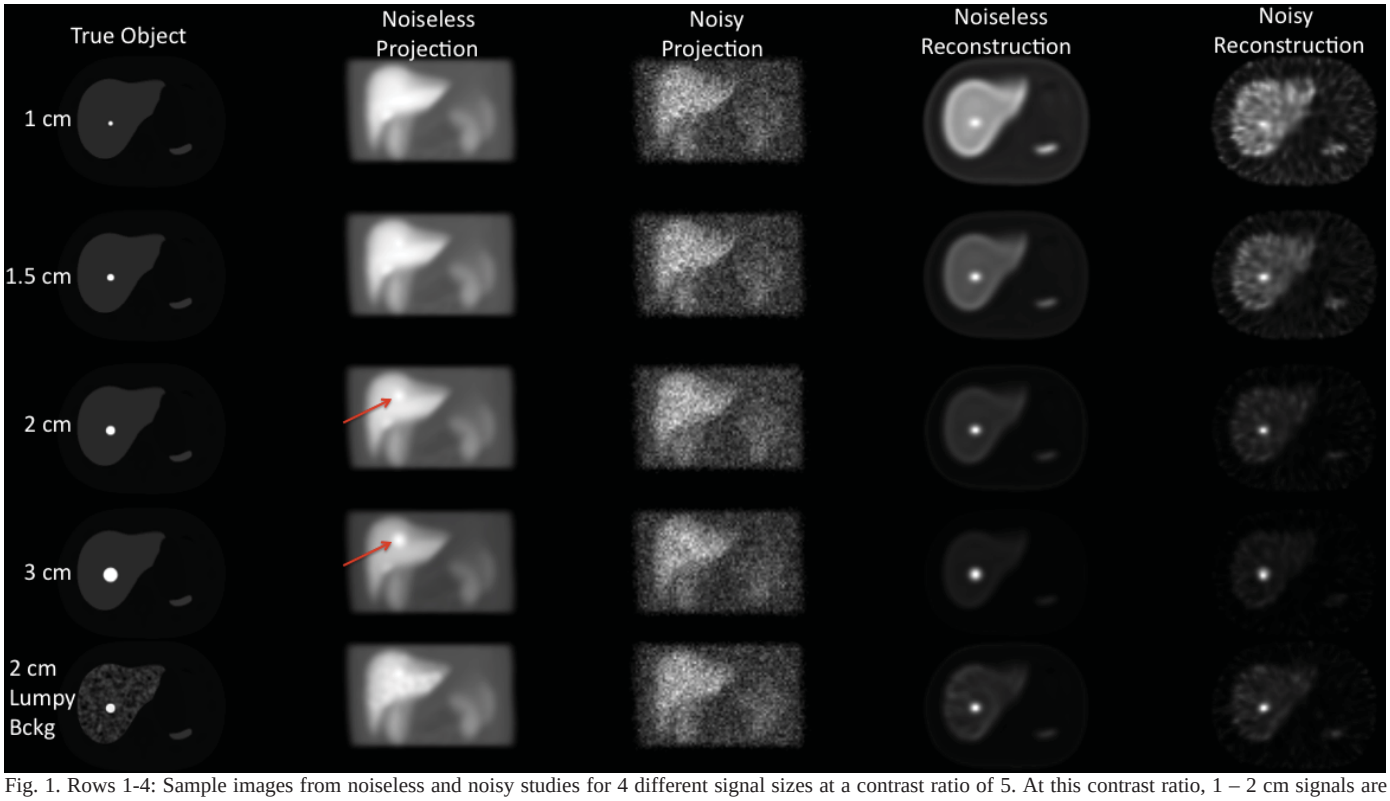


Fig. 1. Rows 1-4: Sample images from noiseless and noisy studies for 4 different signal sizes at a contrast ratio of 5. At this contrast ratio, 1 – 2 cm signals are difficult to identify even in the noiseless projections (anterior view is shown). Last row shows the images for the lumpy background for signal size of 2 cm and contrast ratio of 5.

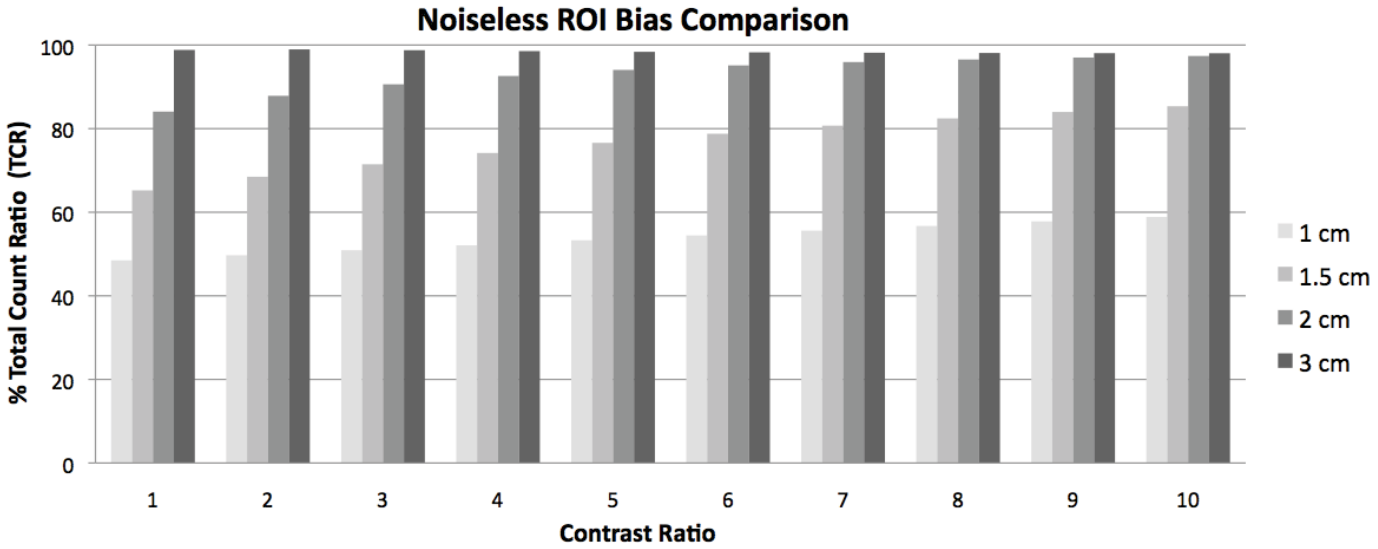


Fig.2. % Total count ratios (TCR) for four signal sizes and ten contrast ratios obtained within the ROI. TCR= 100 x total (estimated) / total (true). The bias of ROI estimation is highly dependent upon both signal contrast and size, which has indications for its utility in monitoring response to treatment. At 2 cm size and relatively high contrast ratio, the bias is small indicated by a TCR close to 100. At 3 cm size, the TCR is close to 100 for all contrast ratios.

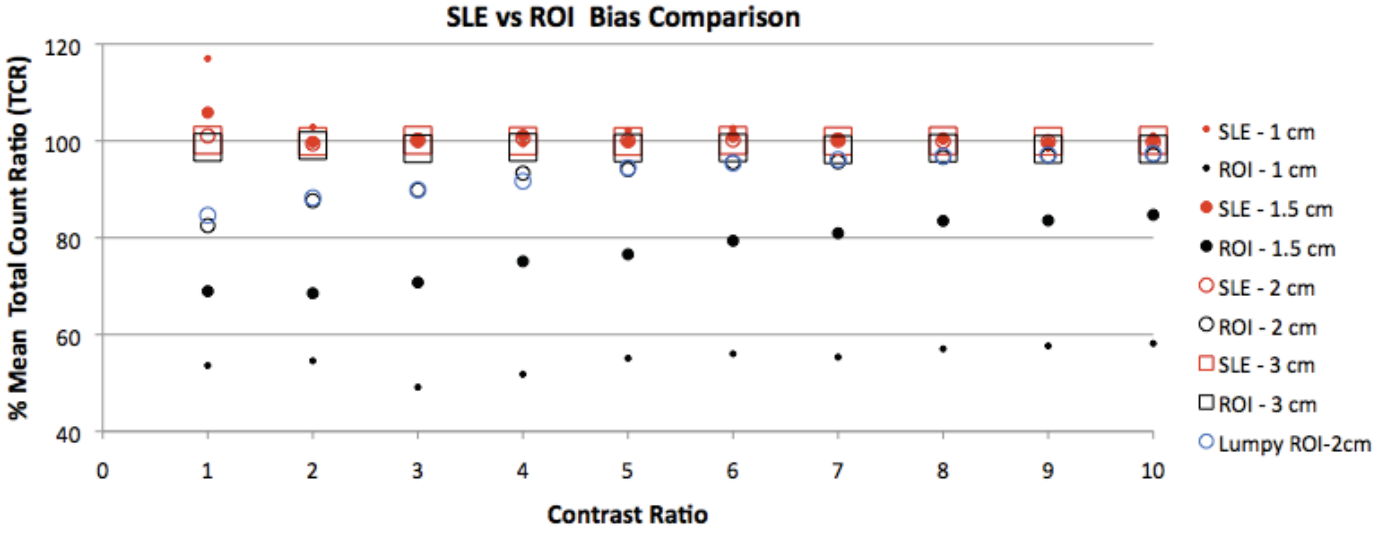


Fig. 3. % Mean total count ratios (TCR) of the SLE (red markers) and ROI (black markers) methods based on 100 noise realizations for each contrast ratio and calculated from TCR= estimated / true total counts. SLE showed unbiased measurements with the mean ratio of ~100 % in all cases except for the lowest contrast ratio for the signals 1 and 1.5 cm diameters. The ROI measurements were biased depending on the signal size and contrast ratio. With increasing signal size and contrast ratio the ROI bias decreased. However, only when the signal size was 3 cm did the mean TCR-ROI value approach 100 % level. Preliminary results for the ROI-lumpy background studies are shown for a signal size of 2 cm (blue markers).

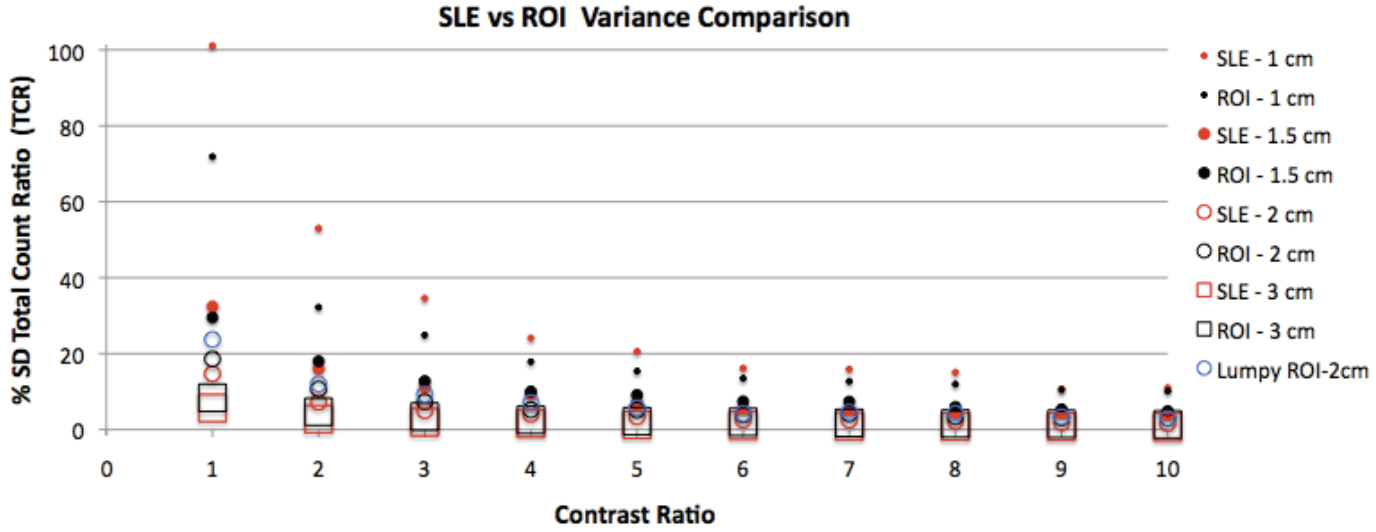


Fig. 4. % Standard deviation of total count ratios (TCR) of the SLE (red markers) and ROI (black markers) calculated from 100 sample images. The SD values decreased with sphere size and contrast ratio for both SLE and ROI methods. SLE showed lower SD for the 1.5, 2, and 3 cm signal diameters for all sphere activity levels except the lowest contrast ratio in 1.5 cm diameter sphere. The variance of the SLE increases for smaller and weaker signals. Preliminary results for the ROI-lumpy background studies for a signal size of 2 cm (blue markers) showed a slight increase in SD compared to the case with constant background.

OpenCL-Accelerated Computation of a 3D SPECT Projection Operator for the Content Adaptive Mesh Model

Francesc Massanes, Jovan G. Brankov
Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL 60616,

Abstract – In this manuscript, we present a preliminary evaluation of a fully 3D projection operator calculation aimed at emission tomography on a non-circular orbit. The proposed methodology uses the content-adaptive mesh model (CAMM) for volumetric data representation. The CAMM is an efficient data representation based on adaptive non-uniform sampling and linear interpolation.

The presented projection operator model incorporates the major data degradation models, namely object attenuation and detector-collimator spatial response, referred to as distance dependent blur. The projection operator is calculated using a ray-casting algorithm and can be adjusted to any scanning geometry and collimator design (e.g. parallel, focusing and pinhole).

Open CL implementation allows shortening of computation time in comparison to standard single CPU implementation.

In this work we successfully tested implementation of the CAMM projection operator by reconstructing images obtained from a realistic data simulation with SIMIND on a non-circular camera orbit. In the future, we will add other collimator designs.

Index Terms – GPGPU, CUDA, OpenCL, Backprojection, Mesh Introduction

I. INTRODUCTION

In tomography it is customary to represent volumetric data using uniform sampling and voxel (pixel) bases functions. In [1] and [2] we proposed a content-adaptive mesh model (CAMM) reconstruction framework in which the image to be reconstructed is represented by non-uniform sampling (see Figure 1) and mesh model basis functions (see Figure 1). In [3] and [4], Sitek et al., proposed a similar adaptive approach under a point cloud name. Both models involve sampling the image domain on a non-uniform grid (mesh nodes), followed by partitioning into a collection of non-overlapping patches (mesh elements).

Figure 1. Emission image (left), attenuation image (center) of NCAT [7] and the mesh structure (right) generated to represent both images.

We reported the initial implementation using graphic processing units (GPUs) acceleration in [5]; however, the evaluation was limited to a circular orbit. In this paper we test an implementation of the projection operator by reconstructing images obtained from a realistic data simulation with SIMIND on a non-circular camera orbit. In the future we will evaluate alternative collimator designs as well.

II. METHODOLOGY

A. Mesh representation of volumetric images

Let f denote a volumetric image function defined over a 3 dimensional (3D) domain $D \subseteq \mathbb{R}^3$ and let us assume that this domain D has already been meshed, so it is divided into a set of M non-overlapping tetrahedrons D_m. Thus, the f function is represented over each D_m as follows:

$$f(\mathbf{x}) = \sum_{n=1}^{4} f\left(\mathbf{x}_n^m\right) \boldsymbol{\varphi}_{n,m}(\mathbf{x}) \quad \text{iff} \quad \mathbf{x} \in D_m$$

where $\varphi_{n,m}(\mathbf{x})$ is the interpolation basis function associated with the n-th node $\mathbf{x}_n$ of D_m. Note that, in the 3D cases, this interpolation basis function is also known as the barycentric coordinates of $\mathbf{x}$ within D_m. This equation can be re-written into vectorial form, leading to the next equation:

$$f(\mathbf{x}) = \varphi_m^T(\mathbf{x}) \mathbf{f}_m \quad \text{iff} \quad \mathbf{x} \in D_m$$

where $\mathbf{f}_m$ and $\boldsymbol{\varphi}_m(\mathbf{x})$ are:

$$\mathbf{f}_m = \left[f\left(\mathbf{x}_1^m\right), f\left(\mathbf{x}_2^m\right), f\left(\mathbf{x}_3^m\right), f\left(\mathbf{x}_4^m\right) \right]^T \text{ and}$$

$$\boldsymbol{\varphi}_m(\mathbf{x}) = \left[\varphi_{1,m}(\mathbf{x}), \varphi_{2,m}(\mathbf{x}), \varphi_{3,m}(\mathbf{x}), \varphi_{4,m}(\mathbf{x}) \right]^T .$$

B. Projection operator calculation by Ray-Casting algorithm

In [2], we showed that the mesh tomographic model can be expressed as follows:

$$E[\mathbf{g}] = \mathbf{H}\,\mathbf{f} + \mathbf{r}$$

where $\mathbf{f}$ denotes a vector formed from the nodal values of the mesh structure representing the emission image (the image which we seek to reconstruct), $E[.]$ is the expectation operator, $\mathbf{r}$ represent random detections, and $\mathbf{g}$ denotes the

collection of the projection bins. Note that the projection matrix, $\mathbf{H}$, depends on the attenuation image and the detector-collimator spatial response as well as the geometry of the tomographic imaging system.

The proposed CAMM projection operator, $\mathbf{H}$, was calculated using a ray-casting algorithm. The ray-casting algorithm projects a ray from each bin into the mesh model. The first step is to find the entry point from the ray into the mesh (lower red point on Figure 2). The algorithm is finished once the ray exits the model (top red point on Figure 2).

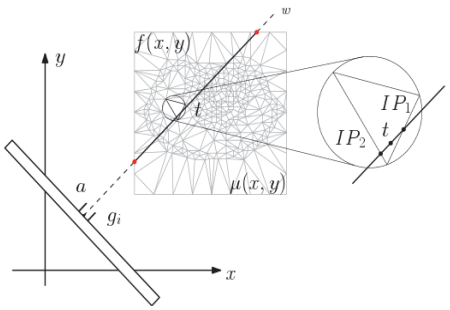


Figure 2. Ray tracing scheme (2D example is shown for clarity).

A parallel implementation on a GPU is straightforward- it requires generating a thread for each ray. Each thread will then have to process a given element *t*, compute how much each vertex contributes to sum along the ray, and compute the exit point of the ray so that it can find the next mesh element to process.

In [5] we also showed a 140 speed up of a GPU based implementation over a convectional central processing unit (CPU) implementation.

Phenomena like attenuation or scattering have been added to the algorithm. Attenuation is computed per each ray, and thus each ray will attenuate its total sum depending on the values of the attenuation map at its path. Scattering is modeled by having more than one ray per bin with slightly different directions and different weighting factor (see [5] for details).

C. CAMM generation

Due to the space constraints of this abstract, we refer the reader to [1] for the specific details on how the CAMM is generated. In short, the CAMM uses a pre-reconstruction as an image to choose the optimal placement of the nodes in the CAMM. This means that some prior knowledge on the reconstructed volume is required in order to properly generate the CAMM. For the pre-reconstructed image, a filtered back projection (FBP) image is typically used.

III. EXPERIMENTS

A. Simulated data

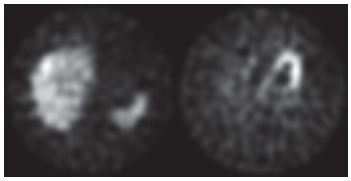
We used the Monte Carlo simulation system SIMIND [9] to simulate a Pricker Prism3000 SPECT system with a low-energy high-resolution (LEHR) collimator and a Tc^{99m} labeled sestamibi as the imaging agent. The emission and attenuation images used have been generated using the 4D NURBS-based cardiac-torso (NCAT) 2.0 phantom [7]. However, since our CAMM model is non-temporal (imaging is not synchronized with cardiac motion) we reduced the 4D phantom to 3D by temporally summing along the time domain.

In order to further demonstrate the ability of our implementation, we chose a non-circular orbit as shown in Fig 3. in which the long radius is 28.5 cm and the short radius is 26.6 cm. We simulated 64 projections.

Figure 3. The orbit used in this experiment.

To test reconstruction performance we chose two different levels of detected photons, the first one having 500,000 counts detected from the heart region and the second simulation, more clinically realistic, has 50,000 detected counts from the heart region.

As we indicated previously, the CAMM requires some prior knowledge to generate the mesh model. To check the correctness of the projection operator first, we used the true attenuation and activity maps (from the phantom) to generate the CAMM which will be used to reconstruct imaging with a high number of counts. However, to fully test the capabilities of CAMM on the 50,000 counts simulation, we used no prior knowledge from the true activity map. Thus, in order to generate the CAMM, we generate pre-reconstruction image by using a filtered back projection (FBP), which assumes circular orbit, with a Butterworth [8] filter of order 4 and cutoff frequency of 0.22 pixels/cycle post reconstruction filtering (see Fig. 4).



Slice 21 Slice 36

Figure 4. FBP images after post reconstruction filtering used for CAMM generation.

B. Hardware and Software

The projection operator was calculated using OpenCL 1.1 in a Tesla C2070 and a sixteen core Intel(R) Xeon(R) CPU E5520 at 2.27Ghz with 94GB of RAM. Since OpenCL can be run equally in both platforms, we split the computation between both of them.

After the projection operator is calculated, we used maximum likelihood expectation maximization (ML-EM) iterative reconstruction [10] using MATLAB in an Intel i7 at 2.4Ghz with 8GB of RAM.

IV. Results

A. High-count results

The first set of simulated images have a 500,000 detected photons from the heart region, the projections are quite noiseless and we have used the true attenuation and activity map to actually generate the CAMM. This can be seen as the best case scenario for the proposed method.

We executed the ML-EM algorithm for 50 iterations and the results are shown in Fig. 5. For comparison purposes we also include the circulate orbit pixel ML-EM and FBP images. Note that pixel ML-EM does not account for the different radius of the camera so, as one can expect, it contains model mismatch artifacts more than the FBP images since FBP is not an iterative method.

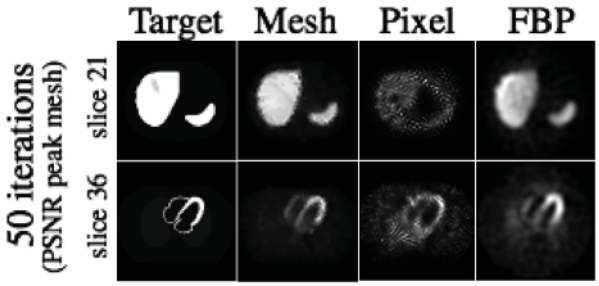


Figure 5. The high count reconstruction results.

B. Low-count results

More practical results are shown next. Here we used noisier data and the CAMM was generated from a pre-reconstructed (using an FBP) image.

These results are shown in Fig. 6 along with a circulate orbit pixel ML-EM. One can observe that the CAMM ML-EM produces the expected results, whereas the pixel ML-EM fails due to the model mismatch.

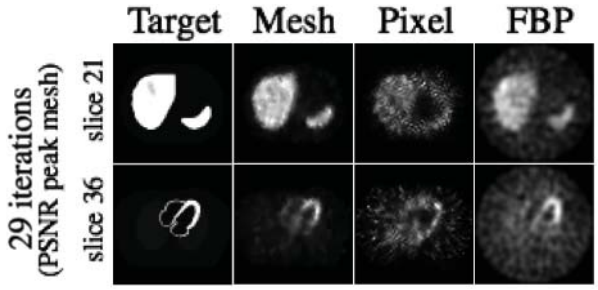


Figure 6. The low count reconstruction results.

Next we evaluated PSNR for the CAMM ML-EM method versus the number of iterations. We also marked PSNR for the FBP images. One can observe that the images presented in Fig. 5 and 6 represent the best-case scenario for CAMM reconstruction.

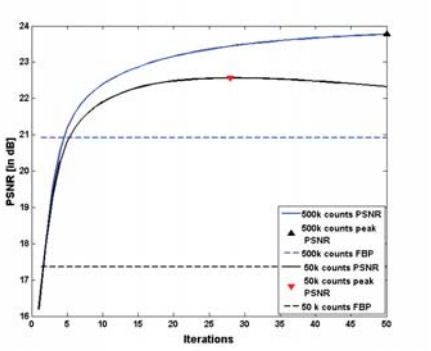


Figure 7. Comparison of the methodologies, the pixel ML-EM will not be shown on this figure.

Finally, in Table 1, we report needed execution time. These results show that proposed CAMM iterative reconstruction has shorter computation time, over pixel based reconstruction, even with overhead of 112 *seconds* needed to compute projection matrix.

V. Conclusions

In this preliminary work we successfully tested implementation of the projection operator by reconstructing images obtained from a realistic data simulation with SIMIND on a non-circular camera orbit. By the time of the conference we hope to expand evaluation section by evaluation heart defect detectability.

Table 1. Computation time

Comp. time	pixel ML-EM		CAMM ML-EM	
	50 iter.	29 iter.	50 iter.	29 iter.
Seconds	335	186	112+24	112+13

VI. References

[1] Y. Yang, M. N. Wernick, and J. G. Brankov, “A fast approach for accurate content-adaptive mesh generation,” IEEE Trans. Image Process., vol. 12, no. 8, pp. 866–881, Aug. 2003.

[2] Jovan G. Brankov, Yongyi Yang and Miles N. Wernick, “Tomographic image reconstruction based on a content-adaptive mesh model”, *IEEE Transactions on Medical Imaging*, vol. 23, pp. 202-212, 2004.

[3] Arkadiusz Sitek, Ronald H. Huesman and Grant T. Gullberg, “Tomographic Reconstruction Using an Adaptative Tetrahedral Mesh Defined by a Point Cloud”, *IEEE Transactions on Medical Imaging*, vol. 25, pp 1172-1179, 2006.

[4] N. F. Pereira and Arkadiusz Sitek, “Evaluation of a 3D point cloud tetrahedral tomographic reconstruction method”, *Phys Med Biol*, vol. 55, pp 5341-5361, 2010.

[5] F. Massanes and J. G. Brankov, “Parallel computation of a SPECT projection operator for a content adaptative

mesh model," Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, pp. 58–61, 2012.

[6] Arkadiusz Sitek, Georges El Fakhri, Jinsong Ouyang and Jonathan S. Maltz. "Fast analytical modeling of Compton scatter using point clouds and graphic processing unit (GPU)", *IEEE Nuclear Science Symposium Conference Record*, vol. 6, pp 4546-4548, 2007

[7] W. P. Segars, "Development and Application of the New Dynamic NURBS-based Cardiac-torso (NCAT) Phantom," University of North Carolina at Chapel Hill, 2001.

[8] S. Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, pp. 536–541, 1930.

[9] M. Ljungberg and S.-E. Strand, "A Monte Carlo program for the simulation of scintillation camera characteristics," *Computer Methods and Programs in Biomedicine*, vol. 29, no. 4, pp. 257–272, Aug. 1989.

[10] M. N. Wernick, *Emission Tomography*, First Edition. Academic Press, 2004, pp. 1–585.

OpenCL Accelerated Multi-GPU Cone-Beam Reconstruction

Marc Käseberg, Steffen Melnik and Erwin Keeve

***Abstract*—Volume reconstruction in cone-beam CT is a computationally demanding task. Since recent years, the reconstruction is accelerated by utilizing Graphics Processing Units (GPUs). Frameworks for General Purpose Computations on GPUs are a proven tool to access the resources of graphics cards. With the Open Computing Language (OpenCL) the first open standard for cross-vendor and cross-platform programming emerged, which allows to accelerate applications in heterogeneous environments.**

In this paper we present an implementation of an OpenCL accelerated volume reconstruction, based on the Feldkamp cone-beam CT algorithm. Our approach enables the utilization of multiple OpenCL devices in parallel. Furthermore the developed data management allows to handle volumes larger than the device memory.

In experiments we proved the portability of our implementation on several devices from different vendors. Additionally, the performance scalability over multiple OpenCL devices was investigated. In a multi-GPU environment consisting of three NVIDIA GTX 580 our approach achieved up to 47.06 Giga Updates per Second and shows a speedup factor of 2.8 over a single GPU reconstruction.

***Index Terms*—GPGPU, FDK, volume reconstruction, OpenCL, heterogeneous systems, cone-beam computed tomography, multi-GPU**

I. Introduction

In recent years, the steadily increasing performance of Graphics Processing Units (GPU) and the advancing development of graphics card programming tools enable the transfer of computationally demanding computations from CPU to GPU. General Purpose Computations on Graphics Processing Units (GPGPU) enable the programmer to perform computations on GPU without any need of knowledge in graphics programming. Proprietary frameworks for GPGPU programming like the Common Unified Device Architecture (CUDA) [1] from NVIDIA are platform and vendor dependent. Hence, programs written in CUDA can only be executed on graphic cards from NVIDIA.

M. Käseberg, S. Melnik and Prof. Dr. Erwin Keeve are with the Fraunhofer Institute for Production Systems and Design Technology IPK, Pascalstrasse 8–9, 10587 Berlin, Germany

Prof. Dr. Erwin Keeve is with the Department of Maxillofacial Surgery and Clinical Navigation, Charité – Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany (e-mail: keeve@charite.de).

However, modern computer systems are heterogeneous systems which do not only consist of a GPU but also of one or more CPUs or additional GPUs. For minimal computation times the full capacity of a system has to be utilized. For that purpose a programming language for heterogeneous systems is needed. With the Open Computing Language (OpenCL) the first open standard for parallel programming of cross-vendor and cross-platform systems emerged. OpenCL [2] enables the use of identical code on different devices like GPUs, CPUs or DSPs. This way computationally demanding tasks can be accelerated independent from the used hardware. Further, OpenCL allows to distribute computations among different devices, e.g. multiple GPUs, what results in further improvements in speed.

In [3] a CUDA based algorithm for GPU accelerated cone-beam reconstruction is presented. By reducing the off-chip memory accesses and memory latency hiding a 512^3-voxel volume with 360 512^2-pixel images was reconstructed in 5.6 seconds. Contrary to the presented approach above, where the entire volume is stored in video memory, the method proposed in [4] allows the reconstruction of volumes which do not fit entirely in video memory. By using the proposed method, volumes up to 2048^3 voxels were reconstructed. Enhanced optimizations were done by Okitsu et al. in [5], where experiments have shown that texture cache optimizations are an important factor. The proposed method organizes threads in a manner, that texture accesses are closely located to maximize cache efficiency. The presented method was evaluated on a single GPU and a multi GPU system consisting of two NVIDIA Tesla C870 graphics cards. However, the volume was stored entirely in video memory. This approach was enhanced in [6] by using up to four graphics cards (NVIDIA Tesla C2070 server). By portioning the volume into smaller subvolumes a 2048^3-voxel volume was reconstructed.

In contrast to the presented CUDA implementations, the cone-beam reconstruction task in [7] was done with OpenCL. For evaluation a volume of the size 256 x 256 x 128 voxels was reconstructed on a single GeForce GTX 280. A performance comparison between CUDA and OpenCL was done in [8] for an optimized FDK implementation. Sawall et al. results have shown, that CUDA outperforms OpenCL. This circumstance was led back to the different development states of OpenCL and CUDA.

II. METHODS

A. FDK

The algorithm of Feldkamp, Davis and Kress [9] is a generalization of the filtered backprojection for cone-beam computed tomography. The FDK solves the reconstruction problem by processing two steps: the pre-weighting and filtering of projection data and the weighted backprojection. The pre-weighting of raw projection data corrects the measured line integrals, since increasing opening angles lead to longer ray lengths from source to detector. After weighting the projection images are row-wise filtered, e.g. with the Ram-Lak filter [10]. Due to performance issues the filtering is done in frequency domain.

The backprojection of an image must be processed after the filtering. During the backprojection the filtered images are projected through every voxel of the volume according to the projection geometries. However, the voxels can be processed in parallel and independent from the order of the images.

B. Implementation of the FDK

In our approach the filtering is done by the CPU whereas the backprojection is performed by one or more OpenCL devices, e.g. GPU. The activity diagram of the implemented FDK algorithm is shown in figure 1. After the filtering of a set of N_s images the images are transferred to the OpenCL device for the backprojection step. In general the filtering of an image on CPU is much faster than its backprojection on an OpenCL device. Therefore, the filtering of the projection data and the backprojection can be performed in parallel.

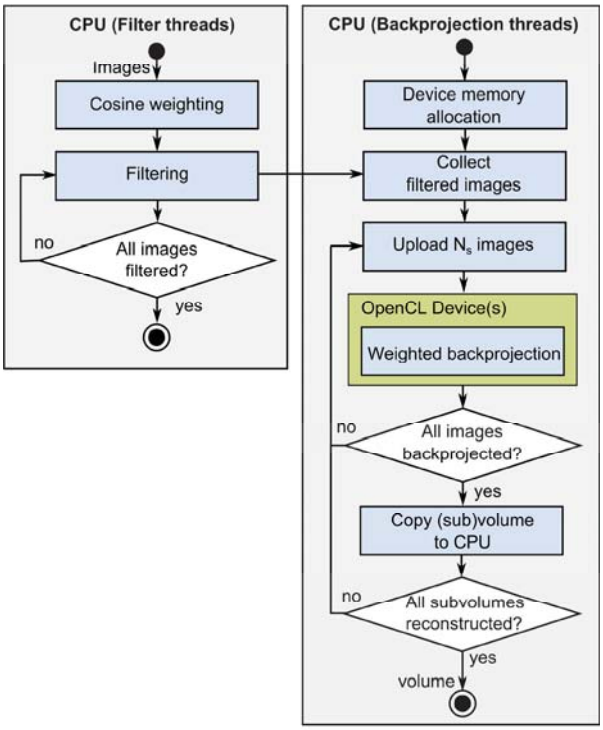


Fig. 1. Activity diagram of the proposed CBCT algorithm for heterogeneous systems.

Compared to the computations the data transfer is often the most time consuming part. Given a filtered 1024^2-pixel image in float32 and an image set size of $N_s = 8$ 32 MB have to be transferred to the OpenCL device. To speed up the algorithm the image data is transferred in parallel to the OpenCL backprojection computations. In this way we practically eliminated the data transfer overhead.

For backprojection a voxel-driven method is used. Since the backprojection is a data parallel task, every voxel is independently processed in a separate OpenCL kernel instance. That way a vast of voxels can be processed in parallel. The backprojection is finished when all projection images were successfully backprojected into the volume.

For reconstruction the volume must be stored in device memory, e.g. video memory. If the volume is that large, that it cannot be stored in device memory, our algorithm divides the volume into smaller subvolumes. The subvolumes are consecutively reconstructed on the device. This approach enables to reconstruct volumes of any size, even if the volume is larger than the available device memory. After reconstruction the (sub)volume is copied to CPU for post processing and visualization. In the case of a reconstruction with multiple subvolumes the parts are merged to a final volume.

C. Kernel Design

During the backprojection the information in every projection image is used to update the volume data. Due to its size the volume is stored in global memory of the OpenCL device. Since global memory accesses are very slow, these accesses are reduced by updating the volume once for N_s images. To benefit from hardware accelerated interpolation on GPU, the projection images are stored in the texture memory. OpenCL 1.2 specifies an array of textures. Unfortunately, the latest NVIDIA OpenCL drivers only support OpenCL 1.1, so that this feature could not be used for our approach and the filtered images are given to the kernel as argument. Since the number of arguments is limited, only a few images can be processed during a kernel execution at a time. To increase the number of images processed during a single kernel execution in [8] multiple images were stored in one texture. For the backprojection of all images the OpenCL kernel is executed multiple times, once for every image set. The implemented backprojection works as follows:

At first the exact position of the voxel center is computed using the global OpenCL thread index. Afterwards a projection matrix is used to calculate the image index where the view ray from the x-ray source position through the current voxel center intersects the image plane [11]. The projection matrix A is pre-calculated for every image on CPU. That way the amount of registers per kernel can be minimized. The image index is calculated according to the following equation where (x, y, z) gives the current voxel position.

$$u = (A_{00}x + A_{01}y + A_{02}z + A_{03}) * w^{-1}$$
$$v = (A_{10}x + A_{11}y + A_{12}z + A_{13}) * w^{-1}$$
$$w = (A_{20}x + A_{21}y + A_{22}z + A_{23})$$

The resulting image indices u and v are given in pixel and *w* is a normalization factor that is also used for backprojection weighting. To get the intensity of the image at the intersection point the four surrounding pixel values are bilinear interpolated. Therefore a hardware accelerated OpenCL in-build function is used. Afterwards the interpolated value is weighted [9] and added to the current voxel value. The calculations are performed without any simplified assumptions about the imaging parameter and acquisition trajectory, e.g. circular path. Since the kernel is executed on various systems, no device specific optimizations were done. Note here, that our approach is also not optimized regarding texture caching yet. Nevertheless, because of significant speedups reported in [12], [13] and [14] the kernel should be improved that way in future work.

D. *Workload balancing*

Since different hardware devices are varying in performance and resources, for multi device execution the reconstruction workload has to be properly distributed among the different devices. The aim is to ensure that every device has been finished with its computation at the same time and that faster devices do not become idle before the reconstruction is done. For that purpose a workload balancing was implemented based on a device specific performance index. Since architectural disparities of different platforms the performance index cannot be derived from hardware properties. Our approach uses a benchmark for performance estimation instead. For an optimal workload balancing the benchmark should correspond with the main application as much as possible. For that reason the implemented backprojection itself is used as benchmark. Due to possible abnormalities in runtime, caused by OpenCL initialization, kernel compilation and memory caching effects, the benchmark is performed multiple times [14]. From the mean value of the computation times the performance index is derived and used to portion the device specific workload.

III. Results and Discussion

A. *Heterogeneous systems*

To evaluate our proposed OpenCL FDK algorithm in respect to cross-vendor and cross-platform capability, we performed the algorithm on different OpenCL devices and computer systems. Additional to single and multiple devices of the same type, the test was performed on a heterogeneous GPU environment consisting of different GPUs from NVIDIA and AMD. For performance comparison only the runtime of the OpenCL backprojection kernel was measured. For all measurements the same kernel was used without any modifications. As test case we performed a reconstruction of a 1024^3-voxel volume with 270 1024^2-pixel projection images on each system. The performance of the tested OpenCL devices, denoted as GUPS (Giga Updates per Second), is shown in figure 2. In our measurements some graphics cards, like the NVIDIA 680 GTX, are slow compared to their theoretical performance. This agrees with the conclusion of [16] and [17], that the OpenCL code is portable but not the performance.

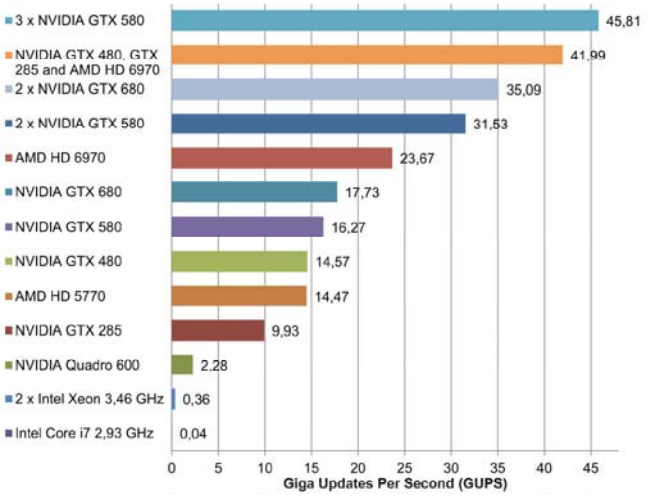


Fig. 2. Evaluation results of our implemented backprojection kernel on different OpenCL devices (GPUs from AMD and NVIDIA and CPUs from Intel).

B. *Multiple GPUs*

In addition to the portability evaluation we have investigated the capability of OpenCL to use multiple devices. For that purpose the FDK algorithm was performed on N = 1, 2 and 3 NVIDIA GTX 580 graphics cards. The measurements were done for a variable volume size from 128^3 up to 2048^3-voxels with 270 images of 512^2, 1024^2 and 2048^2-pixels. In figure 3 the OpenCL backprojection performance in GUPS is shown.

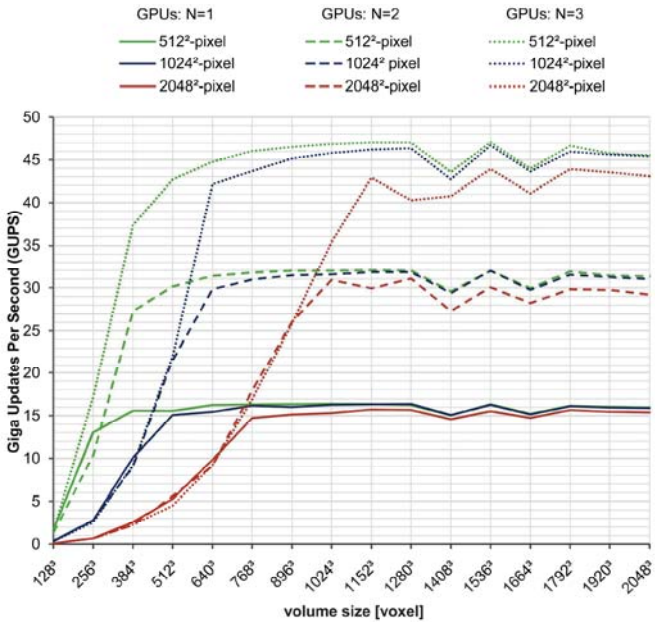


Fig. 3. Backprojection performance for one, two and three NVIDIA GTX 580 graphics cards in Giga Updates per Second (GUPS). For three different image resolutions (512^2, 1024^2 and 2048^2-pixel) the volume size was varied from 128^3 up to 2048^3 voxels.

If one regards the three different characteristic lines of one GPU configuration (N = 1, 2 or 3), it is apparent that the lines for larger volumes nearly match. This results from the overlapping of data transfer and computations. The image resolution only effects the backprojection times of smaller volume sizes, where data transfer takes longer than the computations. For multi-GPU reconstruction this effect appears more early, since the volume is portioned in smaller workloads and the volume size per GPU decreases. The amount of data transfer can be reduced by storing the projection images in half-float precision format. The impact in means of image quality was evaluated in [18]. Another method is to identify the image sections which are actually required for the backprojection computation of the current (sub)volume and only transfer this data [4]. It can also be observed that the vertical gaps between the lines of a GPU configuration (N = 1, 2 or 3) become larger with increasing number of used GPUs. We assume that this behavior is caused by caching effects during bilinear interpolation.

Another characteristic feature of the plots is the performance breakdown for certain volume sizes. For 1536^3 voxels the most GUPs can be observed, while at a volume size of e.g. 1664^3 voxels the performance decreases. Although the number of threads is dynamically adapted, due to hardware restrictions small variations in device occupancy can occur.

On one GPU a reconstruction of the 2048^3-voxel volume with 270 2048^2-pixel projection images takes 142.9 seconds (15.93 GUPS), with two GPUs 75.2 seconds (29.21 GUPS) and about 53.8 seconds (43.06 GUPS) with three graphics cards. This is equal to a speedup of 1.9 and 2.8 compared to the reconstruction with a single GPU. The reached speedup is almost ideal. The not exact linear scaling with the number of GPUs can be traced back to the increasing data management and data transfer overhead.

IV. Conclusion

In this paper we presented an OpenCL implementation of the FDK reconstruction algorithm for CBCT. The algorithm is capable of reconstructing large volumes of any size as long as the volume can be hold entirely in RAM. Further our approach is able to use multiple OpenCL devices in parallel.

The experiments have shown, that OpenCL is well suited for the hardware acceleration of computationally demanding programming tasks on heterogeneous systems. The OpenCL code was performed on every tested device without any modifications. However, the quality and performance of the different vendor implementations of OpenCL are varying. Furthermore, in our experiments OpenCL was able to manage different devices in parallel. During the multi-GPU performance evaluation we achieved a speedup of 1.9 with two graphics cards compared to a single NVIDA GTX 580 and a speedup factor of 2.8 with three GPUs. In future work, we plan to reduce the data transfer between main memory and the OpenCL devices to increase the performance for smaller volume sizes. Also an optimization regarding texture caching seems to be a promising approach for an overall speedup.

References

[1] NVIDIA, "CUDA C Programming Guide 4.2", NVIDIA, 2010, Available: http://docs.NVIDIA.com/cuda/cuda-c-programming-guide/

[2] A. Munshi, B. R. Gaster, T. G. Mattson, J. Fung and D. Ginsburg, "OpenCL Programming Guide", 1st ed., Amsterdam: Addison-Wesley Longman, 2011.

[3] Y. Okitsu, F. Ino and K. Hagihara, "Accelerating cone beam reconstruction using the CUDA-enabled GPU", *High Performance Computing*, HiPC, 2008.

[4] X. Zhao, J. Hu and P. Zhang, "GPU-Based 3D Cone-Beam CT Image Reconstruction for Large Data Volume", *International Journal of Biomedical Imaging*, vol. 2009, 2009, Available: http://www.hindawi.com/journals/ijbi/2009/149079/

[5] Y. Okitsu, F. Ino and K. Hagihara, "High-Performance Cone Beam Reconstruction Using CUDA Compatible GPUs", *Parallel Computing 36(2-3)*, pp. 129-141, 2010.

[6] F. Ino, Y. Okitsu, T. Kishi, S. Ohnishi and K. Hagihara, "Out-of-core cone beam reconstruction using multiple GPUS", In Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, pp. 792-795, 2010.

[7] B. Wang, L. Zhu, K. Jia and J. Zheng, „Accelerated cone beam CT reconstruction based on OpenCL" International Conference on Image Analysis and Signal Processing, IASP, pp. 291–295, 2010.

[8] S. Sawall, L. Ritschl, M. Knaup and M. Kachelrieß, "Performance Comparison of OpenCL and CUDA by Benchmarking an Optimized Perspective Backprojection", *Fully Three-Dimensional - Image Reconstruction in Radiology and Nuclear Medicine*, pp. 12–15, 2011.

[9] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm", *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, Volume 1, Issue 6, pp.612-619, June 1984.

[10] A. C. Kak and M. Slaney, "Principles of Computerized Tomographic Imaging", *IEEE Press*, 1988.

[11] Kachelriess, M., Knaup, M., & Bockenbach, O. (2006). „Hyperfast Perspective Cone--Beam Backprojection". *2006 IEEE Nuclear Science Symposium Conference Record*, 1679–1683. doi:10.1109/NSSMIC.2006.354221

[12] Z. Zheng and K. Mueller, "Cache-aware GPU memory scheduling scheme for CT back-projection", *Nuclear Science Symposium Conference Record,* NSS/MI, pp. 2248–2251, 2010.

[13] E. Papenhausen, Z. Zheng and K. Mueller, "GPU-Accelerated Back-Projection Revisited: Squeezing Performance by Careful Tuning", *Workshop on High Performance Image Reconstruction (Fully 3D Image Reconstruction in Radiology and Nuclear Medicine)*, HPIR, pp. 19-22, July 2011.

[14] H. Scherl, B. Keck, M. Kowarschik and J. Hornegger, "Fast GPU-Based CT Reconstruction using the Common Unified Device Architecture (CUDA)." 2007 IEEE Nuclear Science Symposium Conference Record, pp. 4464–4466, 2007.

[15] V. Podlozhnyuk, (June, 2007), "Image Convolution with CUDA", NVIDIA Corporation, USA Available: http://developer.download. NVIDIA.com/compute/cuda/1.1-Beta/x86_64_website/projects/ convolutionSeparable/doc/convolutionSeparable.pdf

[16] S. Rul, H. Vandierendonck, J. D'Haene and K. De. Bosschere, "An experimental study on performance portability of OpenCL kernels.", *Symposium on Application Accelerators in High Performance,* SAAHPC, pp. 1–3, 2010.

[17] K. Komatsu, K. Sato and Y. Arai, "Evaluating performance and portability of OpenCL programs", *Fifth International Workshop on Automatic Performance Tuning*, 2010.

[18] C. Maaß, S. Steckmann, B. Baer, L. Hillebrand, W. A. Kalender and M. Kachelrieß, "CT image reconstruction with half precision floating point values", *Proceedings of the 2nd Workshop on High Performance Image Reconstruction and the 10th International Meeting on Fully 3D Image Reconstruction*, pp. 24–28, 2009.

Aortic Root Motion Correction in C-Arm Flat-Detector CT

Julia Wicklein, Günter Lauritsch, Kerstin Müller, Holger Kunze, Willi A. Kalender, and Yiannis Kyriakou

***Abstract*—Treatment of cardiac diseases via minimally invasive procedures is of major interest in the clinics. An angiographic C-arm CT system is state-of-the-art in an interventional cardiac laboratory. It opens up the possibility of 3D reconstruction during the procedure. Due to the long acquisition time of several seconds of the C-arm, imaging of dynamic structures is a challenging problem. Therefore, motion correction for cardiac applications is an issue for this imaging device. New minimally invasive procedures like the recently introduced TAVI (transcatheter aortic valve implantation) suffer from cardiac motion. The 3D image of the aorta is acquired during rapid pacing of the patient to minimize the cardiac motion and to reduce the blood flow. We present a new algorithmic approach for motion compensation of the aortic root for TAVI procedures under sinus rythm to make rapid pacing unnecessary. Our optimization routine was tested on three clinical datasets of the aortic root, wherein all three show promising results.**

***Index Terms*—Flat-Detector CT, Cardiac imaging, Aortic root, Motion correction.**

I. Introduction

Transcatheter aortic valve implantation (TAVI) is a minimally invasive procedure that spares high risk or elderly patients open-heart surgeries to treat severe aortic valve stenosis (see Ref. [1]). By default preoperative surgical planning is performed using 3D computed tomography (CT) images. For example the diameter of the annulus of the aortic outflow tract is measured to make the right choice for the prosthetic valve size. Modern hybrid operating rooms are equipped with fixed C-arm systems, providing the physicians with real time 2D fluoroscopic images for guidance during the surgery. Recently, the authors of Ref. [2] introduced an automatic aorta segmentation approach for TAVI.

Ref. [3] introduces a TAVI imaging procedure which makes use of the C-arm CT for both, the 3D volume and the 2D fluoroscopic images. This allows for an accurate and straightforward 2D/3D overlay during the intervention. The short interval between the 3D aquisition and the valve deployment leeds to a better reflection of the patient's anatomy during the intervention. The 3D image is aquired during a 5 seconds scan taking 248 projections over 200°. Selective contrasting was proposed. A pigtail catheter is placed in one of the cusps (typically the non-coronary cusp). Thus, only the aortic root is contrasted. Unfortunately we do not see the aortic outflow tract since this object is beyond the aortic valves. In order to minimize artifacts caused by cardiac motion, rapid ventricular pacing is applied of 180 - 220 bpm and patient breathing is supended. Figure 1 shows an aortic root segmentation illustrating the basic anatomy of a normal aortic valve with three cusps and the ostium of the two coronary arteries.

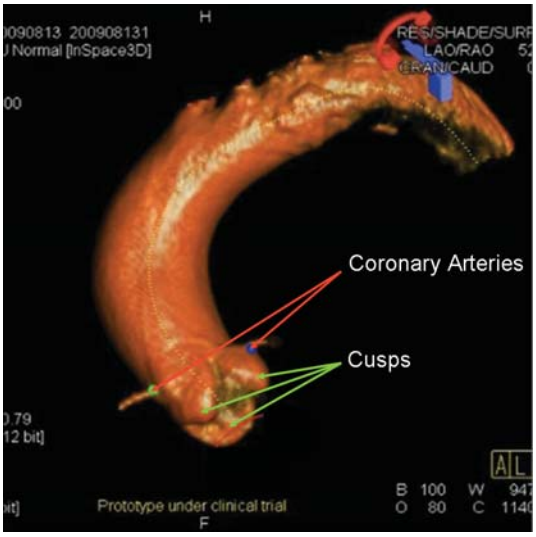


Fig. 1. 3D C-arm CT reconstruction of the aortic root with the TAVI product software from Siemens AG showing typical anatomic landmarks.

Rapid pacing might cause instabilities to the circuit of the patient. There is a clinical desire avoiding rapid pacing and performing imaging in the Sinus rhythm of the patient. Cardiac motion can be treated algorithmically by estimation of the motion from the imaging data and compensating the motion in the reconstruction step. Ref. [4] performs motion estimation on segmented projection images. The accuracy of segmentation is sensitive to the quality and the contrast to noise ratio of the images. Ref. [5] estimates the motion by 3D-3D registration of ECG-gated volume images. A long acquisition time of many heart cycles is needed in Flat Detector CT (FD-CT) for generating ECG-gated volume images (see Ref. [6]). In this paper, we present a novel algorithmic approach to reduce motion artifacts of the aortic root without the need for rapid pacing. It makes use of an entropy based motion and misalignment correction method introduced in Ref. [7].

Julia Wicklein and Willi A. Kalender are with the Institute of Medical Physics (IMP), University of Erlangen–Nürnberg, Henkestraße 91, 91052 Erlangen. Günter Lauritsch, Holger Kunze and Yiannis Kyriakou are with Siemens AG, Healthcare Sector, Forchheim, Germany. Kerstin Müller is with the Pattern Recognition Lab, Department of Computer Science and the Erlangen Graduate School in Advanced Optical Technologies (SAOT), University of Erlangen–Nürnberg, Erlangen, Germany. Corresponding author: Julia Wicklein, E-mail: julia.wicklein@imp.uni-erlangen.de.

The method was originally developed to reduce misalignment and motion artifacts for neuroradiology applications. We adopt the main part of the original approach that is responsible for motion correction tasks. In a two step procedure, aortic root motion is estimated by counteracting variations of the system geometry parameters, which are illustrated in Figure 2.

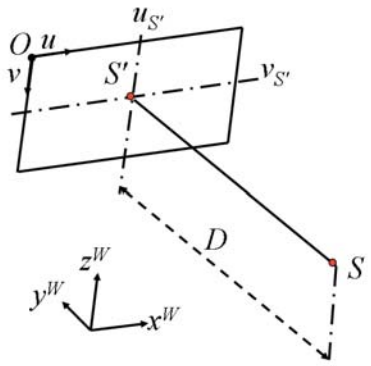


Fig. 2. FD-CT geometry according to Ref. [8]. O defines the detector origin point, S the source position, D the source-detector distance, u, v the detector coordinates and S' the projection of S onto the detector-plane. The rotation axis is parallel to the z^W-coordinate.

The efficiency of our approach was evaluated on three clinical datasets. The data was acquired with the TAVI protocol without rapid pacing. We present the results and illustrate the amenities of our motion correction method for interventional aortic root imaging in the interventional suite.

II. Method

A. Motion Artifacts Metric

In Ref. [8] several image features were investigated with respect to their sensitivity for misalignment artifacts. An entropy criterion based on the gray-level histogram of the reconstructed images was identified to be the most promising one for medical FD-CT applications. Fig. 2 shows the defined FD-CT geometry. Therefore, we chose this feature as motion artifacts metric (MAM) for our motion compensation approach. The histogram (H) of the intensity values q provides a global description of an image. Entropy E using the gray level histogram H is calculated according to:

$$E = -\sum_{q=0}^{Q}(h(q)\cdot log\ h(q)),$$

with

$$h(q) = \frac{H(q)}{N},$$

where Q is the maximum intensity value, h is the normalized histogram or probability distribution of the image and N is the number of image pixels.

B. Optimization Routine

We assume that the motion of the aorta is rigid without major deformations. Therefore, an adaption of the algorithm introduced in Ref. [7] is used to correct for cardiac motion (i.e. orgen- or respiratory motion) without a-priori information. Figure 3 illustrates the estimated parameters which are explained in the following paragraph.

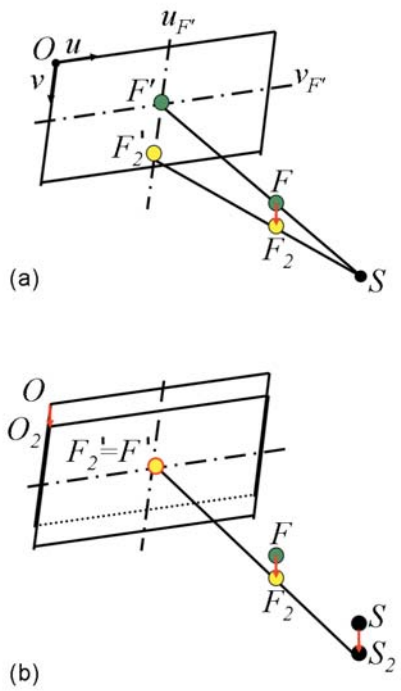


Fig. 3. Example for an object motion: The object point F moves to F_2 (translation in detector-u direction). This causes a translation of the projection of F onto the detector-plane from F' to F_2' (a). Compensation with simultanious detector and source translation (from O to O_2 and from S to S_2) (b).

The movement of the object is compensated by an appropriate variation of the underlying system geometry. A number of four system parameters need to be estimated for motion correction. Parameters like detector- or source-translation or a detector rotation are used to compensate patient motion. Figure 3(a) shows an object translation in detector-u direction. Figure 3(b) shows the compensation by translating detector and source in the same direction. Furthermore a translation in detector v-direction and a detector rotation is optimized to correct object motion in 3D-space.

During the optimization routine, the MAM criterion is used to estimate the mentioned parameters. This is done by minimizing the entropy of the reconstructed images within a gradient descent algorithm with adaptive step size, based on Newton's method:

$$x^{k+1} := x^k + \alpha^k d^k,$$

with

$$d^k = f''(x^k)^{-1}(-f'(x^k)),$$

for updating the function value x^k from iteration k to $k+1$, where d^k defines the Newton-direction with a constant α. The secant method is used to approximate the two derivatives $f'(x^k)$ and $f''(x^k)$ of the optimization function f, representing the MAM criterion:

$$f'(x^k) \approx \frac{f(x^k+1) - f(x^k-1)}{(x^k+1) - (x^k-1)},$$

$$f''(x^k) \approx \frac{f(x^k+1) + f(x^k-1) - 2f(x^k)}{(\frac{(x^k+1)-(x^k-1)}{2})^2}.$$

Entropy minimization is performed during a blockwise and a projectionwise parameter optimization. We reconstructed three transverse slices (z = 0, z = 20 pixels, z = -20 pixels) for the optimization algorithm with a size of 256 pixels $\times$ 256 pixels and a pixel size of 0.5 mm $\times$ 0.5 mm.

1) Blockwise Optimization: The first step iteratively adjusts blocks of projections covering a certain range of the scan. This range enables the correction for cardiac without conflicts of different heart phases. These are adjusted through a number of iterations.

2) Projectionwise Optimization: The second step performs projectionswise adjustments of the whole scan within a few iterations. This procedure removes streak artifacts caused by deviations of single projections.

III. Measurements

We evaluated three datasets scanned with 248 projections over 200° using the Siemens Artis zeego system (Siemens AG, Healthcare Sector, Forchheim, Germany) with a detector of size 616 pixels $\times$ 480 pixels and a pixel size of 0.616 mm $\times$ 0.616 mm, a source-isocenter distance of 785 mm and a source-detector distance of 1200 mm. The aortic root was contrasted with a pigtail catheter placed close to the aortic valve.

IV. Results

The proposed algorithm for motion compensation of the aortic root was evaluated on three clinical datasets. Figures 4, 5 and 6 show the preliminary results. The original standard FDK reconstruction according to Ref. [9] is displayed in (a). It is visible that the image quality of the FDK reconstruction is degraded by motion artifacts. The three leaflets are highly corrupted by motion blur. The motion corrected reconstructions in Figure 5(b), 6(b) and 7(b) show the improvements in the area of the aortic root. Each Subfigure shows the multi-planar reconstruction images (long axis view top left and right, short axis view bottom left) and volume rendering (bottom right). The results presented in Figure 4, 5 and 6 demonstrate the effect of the optimization approach. The aortic roots after optimization appear less artifact afflicted. The cusps of the aortic valve and even the commissures of the leaflets are specifiable after applying the correction routine. We tested the approach on three different datasets, wherein all three show comparable good results.

V. Conclusions

The first results with the new motion correction algorithm on three clinical datasets show very promising results. The motion correction approach works without a-priori knowledge and gives the possibility to do interventional aortic root imaging without the need for rapid pacing or with less contrast agent insertion.

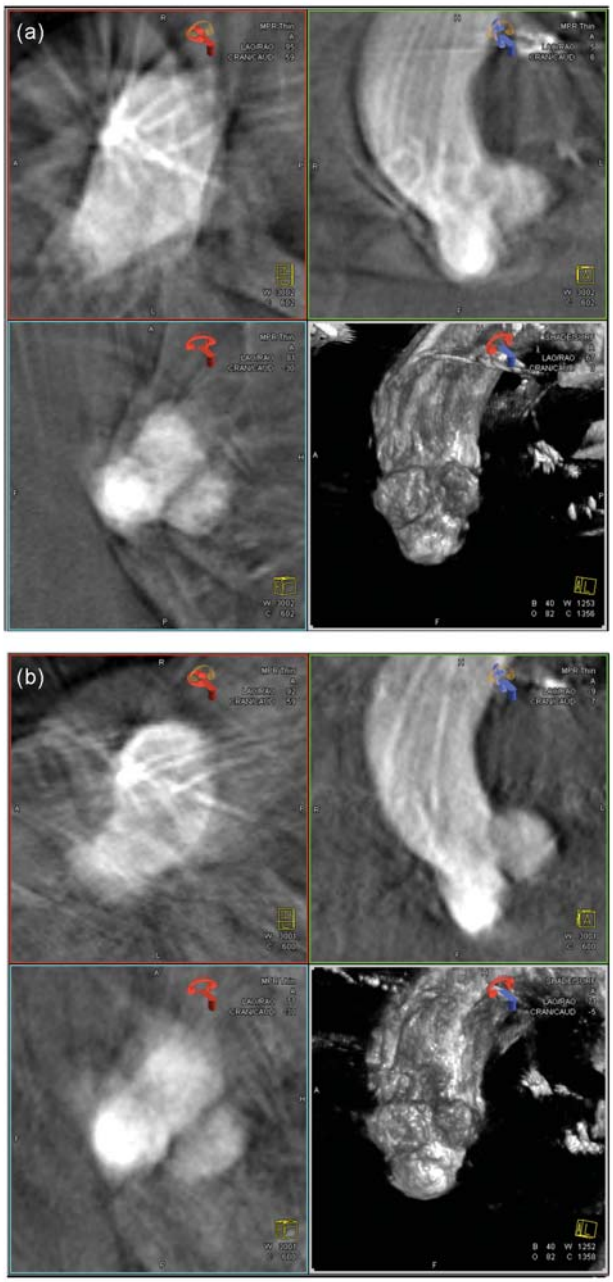


Fig. 4. Aortic root reconstruction of dataset 1: Original (a). Optimization result (b). Each including the saggital cut ($x = 0$, upper left), the coronal cut ($y = 0$, upper right), the transverse cut ($z = 0$, lower left) and the volume image (lower right).

VI. Acknowledgement

The authors would like to thank Prof. James N. Slater of the NYU Langone Medical Center for providing clinical datasets.

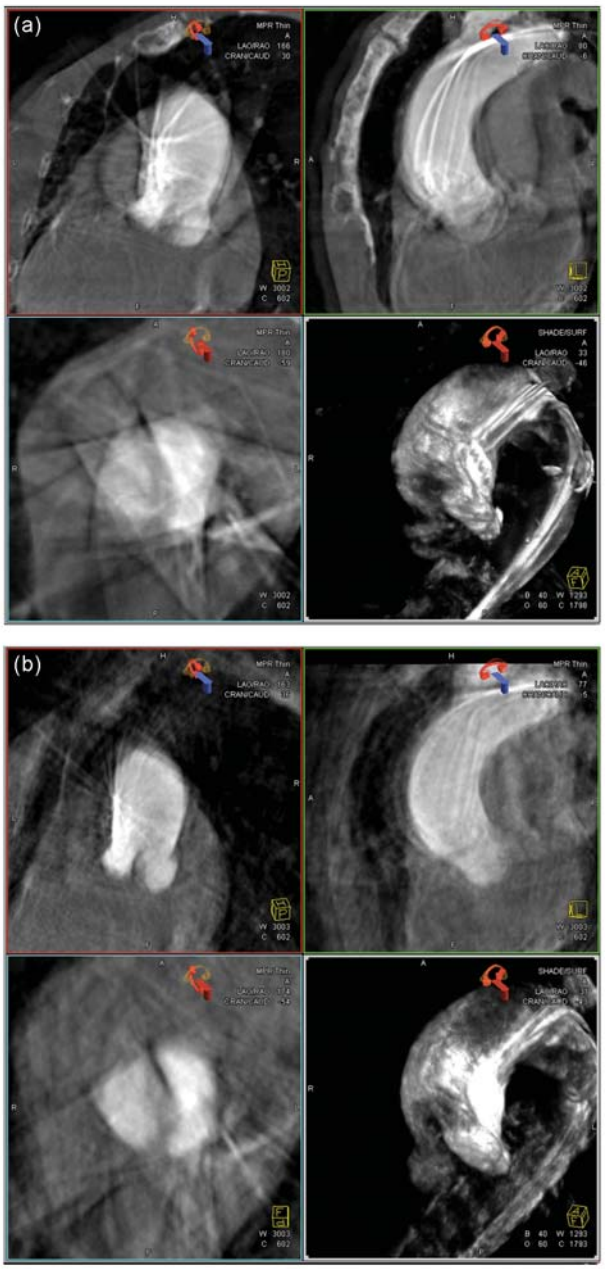


Fig. 5. Aortic root reconstruction of dataset 2: Original (a). Optimization result (b). Each including the saggital cut ($x = 0$, upper left), the coronal cut ($y = 0$, upper right), the transverse cut ($z = 0$, lower left) and the volume image (lower right).

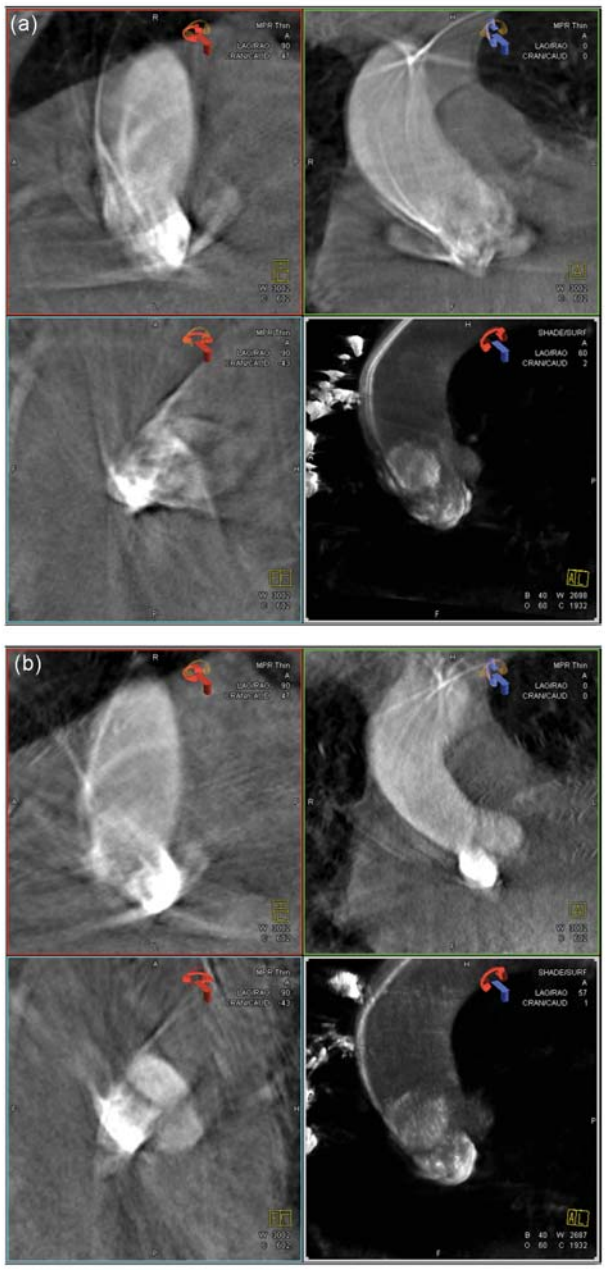


Fig. 6. Aortic root reconstruction of dataset 3: Original (a). Optimization result (b). Each including the saggital cut ($x = 0$, upper left), the coronal cut ($y = 0$, upper right), the transverse cut ($z = 0$, lower left) and the volume image (lower right).

The concepts and information presented in this paper are based on research and are not commercially available.

REFERENCES

[1] M. B. Leon, et. al., *Transcatheter Aortic-Valve Implantation for Aortic Stenosis in Patients Who Cannot Indergo Surgery*, N Engl J Med 2010, No. 363, pp. 1597-1607, 2010.

[2] Y. Zheng, et. al., *Automatic Aorta Segmentation and Valve Landmark Detection in C-Arm CT for Transcatheter Aortic Valve Implantation*, IEEE Transactions on Medical Imaging, Vol. 31, No. 12, pp. 2307-2321, 2012.

[3] M. John, et. al., *System to Guide Transcatheter Aortic Valve Implantations Based on Interventional C-Arm CT Imaging*, Proc. of MICCAI 2010, Part I, LNCS 6361, pp. 375-382, 2010.

[4] K. Müller, et. al., Med. Phys. 2013, to be published.

[5] K. Müller, et. al., *4-D Motion Field Estimation by Combined Multiple Heart Phase Registration (CMHPR) for Cardiac C-arm Data*, Proc. of IEEE NSS-MIC, pp. 3707-3712, 2012.

[6] S. de Buck, et. al., *A new approach for prospective gated cardiac rotational angiography*Proc. of SPIE Medical Imaging, paper 8668-104, 2013.

[7] J. Wicklein, et. al., *An online motion- and misalignment-correction method for medical flat-detector CT*, Proc. of SPIE Medical Imaging, paper 8668-62, 2013.

[8] J. Wicklein, et. al., *Image features for misalignment correction in medical flat-detector CT*, Medical Physics, Vol. 39(8), pp. 4918-4931, 2012.

[9] L. A. Feldkamp, et. al., *Practical Cone-Beam Algorithm*, JOSA A, Vol. 1, Issue 6, pp. 612-619, 1984.

Total Variation Regularization in Digital Breast Tomosynthesis: Regularization Parameter Determination based on Small Structures Segmentation Rates

U. Heil*, S. Fränkel*†, K. Wunder†, D. Groß*, R. Schulze‡, U. Schwanecke§, C. Düber†, E. Schömer*, and O. Weinheimer¶

*Institute of Computer Science
Johannes Gutenberg-University Mainz (JGU), Staudingerweg 9, 55128 Mainz, Germany
Email: heilu@uni-mainz.de

†Department of Radiology, University Medical Center (UMC) of the JGU, 55131, Germany

‡Department of Oral Surgery (and Oral Radiology), UMC of the JGU, 55131, Germany

§Department of Design, Computer Science and Media, RheinMain University of Applied Sciences, 65197, Germany

¶Department of Diagnostic and Interventional Radiology, University Hospital of Heidelberg, 69120, Germany

***Abstract*—Regularization approaches for the limited-angle reconstruction problem in digital breast tomosynthesis are widely-used. Though, their benefits depend largely upon a suitable regularization parameter estimation. We aim to evaluate the reconstruction quality of precise small contrast features objectively with the help of an automated process. These features were represented by so-called Landolt ring (LR) structures of descending sizes contained in an especially designed mammography test object (Quart Mam/Digi Phantom).**

A GPU-based iterative Barzilai-Borwein (BB) algorithm is applied to solve the inverse reconstruction problem using total variation (TV) regularization. Exemplarily, we analyzed limited-angle breast projection images from a commercially available digital breast tomosynthesis (DBT) system (Siemens Mammomat Inspiration). We show that the TV regularization parameter and number of iterations can be chosen in such a way that the detection rate for the LR features is considerably higher than that obtained from the manufacturer's reconstruction (modified filtered backprojection).

***Index Terms*—X-ray tomography, computed tomography, reconstruction algorithms, iterative algorithms, mammography, digital breast tomosynthesis, cone-beam geometry.**

I. Introduction

Breast cancer remains a significant threat to woman's health and the earlier the detection, the higher the chances for good healing prognoses. As tumor size at diagnosis is one of the main predictive factors for survival, all efforts are made to improve detection of small lesions. The 2D mammography is still the standard diagnostic method for screening and the diagnostic setting, although many studies showed limited sensitivity in dense breast tissue [1]. Now, that fast detectors and computers are available, standard tomography has been revitalized in breast diagnostics. In this technique, a 3D volume, respectively a stack of 2D slices, is computed by the use of a few projected X-ray images. The generated 3D information in digital breast tomosynthesis (DBT) should improve lesion detection through reduction of superimposition. For reconstruction of the volume there are different algorithms like filtered back-projection (FBP), shift-and-add (SAA) or algebraic reconstruction techniques (ART). The results in [2] indicate that there may also be a substantial advantage in using TV regularization for microcalcification imaging. In [3] the influence of TV regularization on digital breast tomosynthesis data taken from a Hologic Selenia Dimensions system was analyzed.

In this paper, we investigate iterative image-reconstruction in DBT based on ART and TV with respect to detection of small clearly defined contrast features e.g. Landolt rings (LR) (see Fig. 1,2).

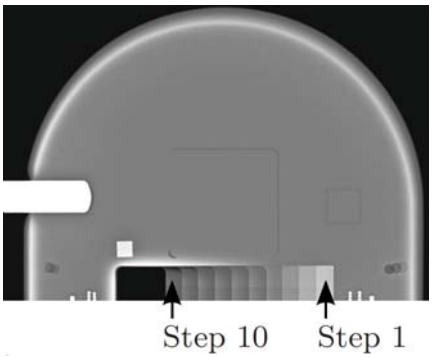


Fig. 1. 2D projection of the Quart Mam/Digi phantom. Step 11 and 12 are not displayed.

ART formulates the projection of a volume to images as the system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{1}$$

where $\mathbf{x} \in \mathbf{R}^n$ is an unknown 3D volume composed of n voxels written as a vector, $\mathbf{y} \in \mathbf{R}^{pm}$ is the set of p 2D images,

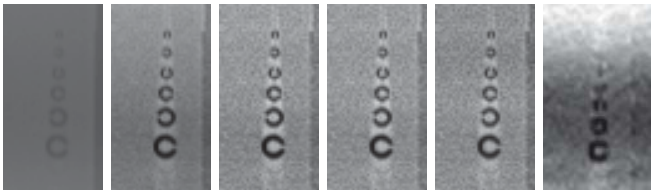


Fig. 2. A reconstructed Landolt ring sequence at one step of the Quart Mam/Digi phantom: six LRs at step 10, 200 mAs, 30 kVp, BB reconstruction, TV $\lambda = 0.2$, iterations 1 up to 20. On the right-hand side the corresponding Siemens Mammomat Inspiration reconstruction at step 10 with 200 mAs is shown.

each consisting of m pixels. The matrix $\mathbf{A} \in \mathbf{R}^{pm \times n}$ is the discretization of a line integral operator defined by the X-ray geometry. Solving (1) means reconstructing the volume. Since this inverse problem is ill-posed, it is also important to consider additional information. Other groups have shown, that TV-regularization leads to a better signal to noise ratio and to a reduction of streaking artifacts [4], [5], but one may ask, whether small structures are better recognizable. With TV-regularization the linear equation becomes a convex optimization problem of the form

$$\min_{\mathbf{x}} f(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda TV(\mathbf{x}) \ . \tag{2}$$

For feasibility reasons, we use a differentiable approximation of the total variation $TV(\mathbf{x})$ which is defined as

$$TV(\mathbf{x}) := \sum_{ijk} \sqrt{d(x_{i,j,k}) + \beta^2} \tag{3}$$

with a small $\beta > 0$ and

$$\begin{aligned} d(x_{i,j,k}) \quad := \quad & (x_{i-1,j,k} - x_{i,j,k})^2 + & (4)\\ & (x_{i,j-1,k} - x_{i,j,k})^2 + & (5)\\ & (x_{i,j,k-1} - x_{i,j,k})^2 & (6) \end{aligned}$$

where the indices i, j and k denote the position in the 3D volume. In [6] the iterative Barzilai-Borwein (BB) solver was successfully used for reconstruction of low-dose cone-beam computed tomography (CBCT) images, delivering good results after just a few (12-30) iterations.

II. Materials and Methods

A. Tomosynthesis System

The system that we used for our reconstructions is a Siemens Mammomat Inspiration. Its X-ray source moves with an angular range of maximal $50\,^\circ$ and under our operating conditions between $-24.82\,^\circ$ and $+21.14\,^\circ$. During this process the system exposes 25 projection images with a size of about 2400×3600 pixels and a pixel spacing of 0.085×0.085 [mm]. The rotation center is at 4.7 cm above the detector and the distance between the x-tube and the detector is about 66 cm [7]. We evaluate one series with 30 peak kilovoltage (kVp) and an overall exposure of 200 mAs (8 mAs per projection), which is higher than the system's automatic mode with an exposure value defined at 105 mAs for 30 kVp for the mammography phantom.

B. Reconstruction Method

The volume, that we reconstruct, is a box with about $2400 \times 3600 \times 47$ voxels and a voxel size of $0.085 \times 0.085 \times 1\,\mathrm{mm}^3$. It is located close upon the detector and contains the whole phantom. To avoid artifacts and minimize the memory required, our volume is defined exclusively by voxels that are projected into regions of the phantom. Exterior voxels are set to 0.

We applied the algorithm by Barzilai and Borwein, which is based on a Quasi-Newton-Method [8]. Thereby an iteration step has the form $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{H}_n^{-1} \nabla f(\mathbf{x}_n)$ where $\mathbf{H}_n$ is an approximation to the Hessian of $f(\mathbf{x})$. Barzilai and Borwein set $\mathbf{H}_n^{-1} = \alpha_n \mathbf{I}$ where α_n is given by

$$\alpha_n = \frac{(\mathbf{x}_n - \mathbf{x}_{n-1})^T (\nabla f(\mathbf{x}_n) - \nabla f(\mathbf{x}_{n-1}))}{(\mathbf{x}_n - \mathbf{x}_{n-1})^T (\mathbf{x}_n - \mathbf{x}_{n-1})} \tag{7}$$

minimizing $\|(\mathbf{x}_k - \mathbf{x}_{k-1}) - \alpha_n (\nabla f(\mathbf{x}_n) - \nabla f(\mathbf{x}_{n-1}))\|$.

Since the total variation is not differentiable as a function of $\mathbf{x}$, we use the differentiable approximation $TV(\mathbf{x})$ as defined in (3). Then, the iteration step becomes

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \alpha_n \left(2\,\mathbf{A}^T (\mathbf{A}\mathbf{x} - \mathbf{y}) + \lambda \nabla (TV(\mathbf{x})) \right) \ . \tag{8}$$

The influence of TV on the reconstruction process can be managed by the regularization constant λ in (2) and (8) respectively. To speed up the reconstruction process forward and back projection ($\mathbf{A}$ and $\mathbf{A}^T$, respectively) are written as shaders running on the graphics processing unit (GPU), see [9].

C. Quart Phantom

To evaluate the quality of our reconstructions we used a new mammography phantom, the Quart Mam/Digi phantom [10]. The most interesting features for our reconstruction are the so-called Landolt rings (see Fig. 1,2). These are special rings with a gap in one of the four directions: right, left, bottom or top. The phantom has 12 steps with increasing densities and each step contains a group of six LRs with diameters from $800\,\mu$m down to $260\,\mu m$.

Fig. 1 shows a projection image of the Quart Mam/Digi phantom. A detailed reconstruction of a group of Landolt rings is depicted in Fig. 2. Furthermore, in the latter the identical region from the Siemens Mammomat Inspiration reconstruction is given. The more LRs are detected correctly in a reconstruction, the better the image quality is. For a fast and objective evaluation we implemented a fully automatic LR detection algorithm based on standardized 12-bit DICOM input datasets.

D. Automatic Landolt Ring Detection

Fig. 3(a) shows a schematic representation of a LR. To measure the detection quality of a LR, three features are calculated (see also Fig. 3(b)):

Fig. 3. (a) LR with gap on the right side. (b) Marked features: center of ring (dot), path inside the ring (line), path in the gap (dotted line) and circle-path outside (line). (c) LRs on Step 7, Exposure 99 mAs, Barzilai Borwein reconstruction, TV $\lambda = 0.5$. (d) Visual output of the automatic LR detection for the first Ring in (c): center, ring and gap are marked correctly. (e) Values on the circle-path through the ring and the gap: the highest peak belongs to the gap.

1) Contrast c, based on the gray value v_1 at the center, the mean value v_2 of the intensities along a circular path on the ring and the mean value v_3 on the circle-path outside: $c = ((v_1 - v_2) + (v_3 - v_2))/2$,
2) Standard deviation sd of the ring values,
3) Difference d between mean gap value and mean ring value.

The calculations are performed with sub-pixel accuracy using bilinear interpolation. The positions of the 12 groups of LRs in the phantom are fix. In order to ensure a more flexible usability of the detection method, offset jumps from an automatically detected landmark to the LR groups are used. Caused by small inaccuracies in the landmark detection, a small search window of $0.5 \times 0.5 \times 1.0\,\mathrm{mm}^3$ for searching the center of the first LR of a group is used - this ensures that the first ring of a group can be determined correctly. A ring is marked at the position where the sum

$$D = \omega_1 c + \omega_2 sd + \omega_3 d, \text{ with } \boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3) = (3, -1, 1) \tag{9}$$

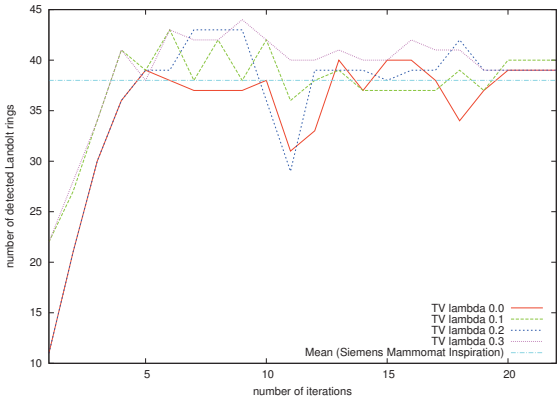
is maximized, varying the center of the LR and the position of the gap. $\boldsymbol{\omega}$ was heuristically defined. A ring is counted as detected if the detection sum D is greater than a threshold κ and the attitude of the gap is correctly detected. Suitable values for the detection threshold κ can be chosen taking the density range of the reconstructed DICOM datasets into account. The correct gap positions are known a priori for all LRs of the phantom.

III. Results

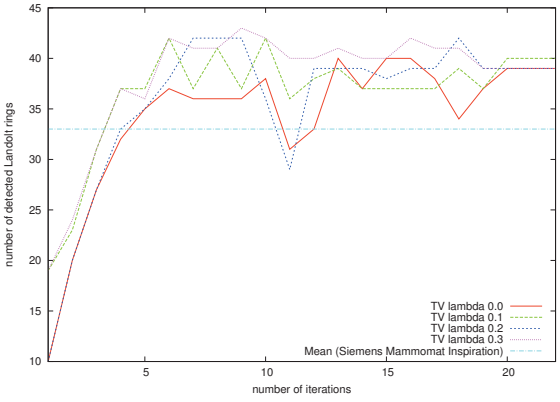
We assign 0.0, 0.1, 0.2, 0.3, 1.0 and 10.0 to the TV regularization constant λ and compare the reconstructions using up to 22 iteration steps. The Landolt ring detection threshold κ is varied to control its influence on the detection rates.

Fig. 4 shows some volume statistics for assigned λ values $(0, 10^{-1}, 10^0, 10^1)$. On the left side (a) the residual norm $||y - Ax||_2$ is plotted versus the iteration number, whereas on the right side the total variation $||x||_{TV}$ of the volume is shown. As one expects, the figure shows decreasing residual norms in (a) and simultaneously increasing total variations of the volumes in (b) with respect to the regularization parameter. Already after 20 iteration steps the desired regularization characteristics are achieved. The differences within the volume statistics for $\lambda \in [0.1, 0.3]$ are marginal. Thus they are omitted in Fig. 4.

Fig. 5 shows the LR detection results for assigned λ values ($\lambda = 0.0, 0.1, 0.2, 0.3$). Two different LR detection thresholds κ were used to define a correct ring count $D > \kappa$ (see (9)): in Fig. 5 (a) $\kappa = 50$ and in (b) $\kappa = 75$. For a higher detection threshold κ less rings were detected e.g. 33 instead of 38 for the Siemens reconstruction. Our results seem to be nonsensitive to threshold variations, because of the higher contrast of our volumes. Both shapes of our data profiles look nearly identical.



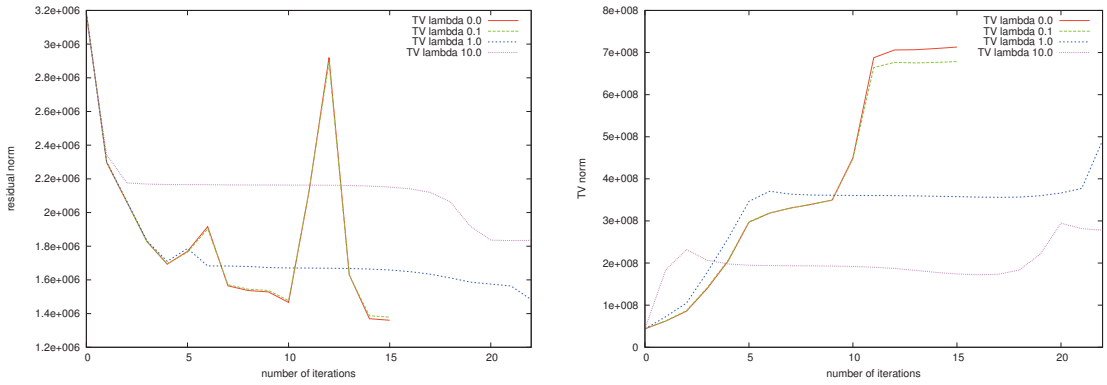
(a) LR detection threshold $\kappa = 50$



(b) LR detection threshold $\kappa = 75$

Fig. 5. LR counts for TV reconstructions with different λ values. In comparison to the Siemens reconstruction $\lambda = 0.1, 0.2, 0.3$ values mostly deliver higher LR counts after 4 iterations.

The computer used for reconstruction had an Intel Core i7 CPU with 2.97 GHz clock speed and 12 GB RAM. We used a 64-bit Windows 7 OS; the GPU is a NVIDIA GeForce



(a) (b)

Fig. 4. Iterative reconstruction characteristics for different TV regularization values $\lambda = 0, 10^{-1}, 10^{0}, 10^{1}$: (a) residual norm $||y - Ax||_2$ and (b) TV norm $||x||_{TV}$.

GTX 280. The reconstruction program is compiled as an 32-bit application. Intermediate reconstruction volumes needed during the reconstruction process had to be stored on the hard disk, because of the large size of the Quart Mam/Digi phantom. The runtime for one iteration was approx. one up to two minutes.

IV. Discussion

We showed in this paper that the Landolt ring component of the specific phantom is an adequate tool for the evaluation of DBT algorithms with respect to the representation of small dense structures. TV regularization yielded better perceptibility of the LRs contained in the Quart Mam/Digi phantom, when the parameter λ is adjusted carefully. With the correct iteration number we find more rings, than the reconstruction provided by the manufacturer. A possible explanation for this is a lower standard deviation sd of the values on the ring and thus less noise for comparable contrast values, when using the non-linear TV-regularization.

The following additional steps are planed as future research. We will systematically evaluate our reconstructions for more than 20 iterations over the whole exposure range and time. Investigation and possibly modification of other regularization methods e.g. the L1 norm are planed. With ongoing improvement of the algorithm the required radiation dose could possibly be further reduced.

We will compare the results of the automatic LR detection with the reading of radiologists to evaluate reconstruction quality in a clinical context. Furthermore, we hope that our method can be applied in the context of limited angle breast tomography in order to improve the detection of clinical pathologies (e.g. microcalcifications).

V. Conclusion

We conclude that our iterative TV-regularized reconstruction method can be almost optimally adapted to improve the depiction of small clearly defined contrast features in limited-angle cone-beam reconstruction problems.

Acknowledgment

The authors would like to thank the company Quart GmbH for giving us the opportunity to work with the Mam/Digi phantom as well as Dr. Jochen Schenk from the Radiology Institute Hohenzollernstraße of Koblenz for making available the Siemens Mammomat Inspiration system.

References

[1] M. A. Helvie, "Digital mammography imaging: breast tomosynthesis and advanced applications," *Radiologic clinics of North America*, vol. 48, no. 5, pp. 917–929, 2010.

[2] E. Y. Sidky, X. Pan, I. S. Reiser, R. M. Nishikawa, R. H. Moore, and D. B. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Medical Physics*, vol. 36, no. 11, p. 4920, 2009.

[3] S. Fränkel, K. Wunder, U. Heil, D. Gross, R. Schulze, U. Schwanecke, C. Düber, E. Schömer, and O. Weinheimer, "Influence of total variation regularization on the representation of small structures in digital breast tomosynthesis - a phantom study," 2013, accepted for "Bildverarbeitung für die Medizin".

[4] I. Kastanis, S. Arridge, A. Stewart, S. Gunn, C. Ullberg, and T. Francke, "3d digital breast tomosynthesis using total variation regularization," in *Proceedings of the 9th international workshop on Digital Mammography*, ser. IWDM '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 621–627.

[5] E. Sidky, Y. Duchin, I. Reiser, C. Ullberg, and X. Pan, "Optimizing algorithm parameters based on a model observer detection task for image reconstruction in digital breast tomosynthesis," in *NSS/MIC, IEEE*, 2011, pp. 4230 – 4232.

[6] J. C. Park, B. Song, J. S. Kim, S. H. Park, H. K. Kim, Z. Liu, T. S. Suh, and W. Y. Song, "Fast compressed sensing-based cbct reconstruction using barzilai-borwein formulation for application to on-line igrt," *Med Phys*, vol. 39, no. 3, pp. 1207–1217, 2012.

[7] T. Mertelmeier, J. Speitel, and C. Frumento, "3d breast tomosynthesis - intelligent technology for clear clinical benefits," Siemens, Henkestrasse 127, DE-91052 Erlangen, Germany, White Paper A91XP-30011-25C1-7600, 2012, also available as http://www.medical.siemens.com/siemens/en_GLOBAL/gg_sps_FBAs/files/apps/be_sure/3D_breast_tomosynthesis-intelligent_technology_or_clear_clinical_benefits.pdf.

[8] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *Journal of Numerical Analysis*, vol. 8, pp. 141–148, 1988.

[9] D. Gross, U. Heil, R. Schulze, E. Schömer, and U. Schwanecke, "Gpu-based volume reconstruction from very few arbitrarily aligned x-ray images," *SIAM J. Scientific Computing*, vol. 31, no. 6, pp. 4204–4221, 2009.

[10] Quart mam/digi website. QUART Medizintechnische Geräte GmbH, Kirchenweg 7, DE-85604 Zorneding, Germany. [Online]. Available: http://quart.de/en/test-phantoms/mammography.html

Curvelet sparse regularization for differential phase-contrast X-ray imaging

Matthias Wieczorek*, Jürgen Frikel†, Jakob Vogel*, Franz Pfeiffer‡, Laurent Demaret†, and Tobias Lasser*†

* Chair for Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany
wieczore@cs.tum.edu

†Institute of Biomathematics and Biometry, Helmholtz Zentrum München, Germany

‡ Chair of Biomedical Physics, Technische Universität München, Germany

***Abstract*—Differential phase contrast imaging (DPCI) enables the visualization of soft tissue contrast using X-rays. In this work we introduce a reconstruction framework based on curvelet expansion and sparse regularization for DPCI. We will show that curvelets provide a suitable data representation for DPCI reconstruction that allows preservation of edges as well as an exact analytic representation of the system matrix. As a first evaluation, we show results using simulated phantom data.**

I. Introduction

One of the main shortcomings of conventional x-ray computed tomography (CT) is the low contrast within the soft tissue regions. Differential phase-contrast imaging (DPCI) is an emerging imaging modality which was developed to address this issue. It was shown in [1], that this technique improves the visualization of soft tissues upon conventional X-ray computed tomography (CT).

The basic idea of DPCI consists in performing phase-sensitive x-ray measurements (rather than imaging the absorption coefficient) and the reconstruction of the refraction coefficient $f : \mathbb{R}^2 \to \mathbb{R}$. The mathematical model describes the relationship between the phase change and the refraction coefficient f in the following way, cf. [1], [2], [3],

$$\mathcal{P}f(\theta, s) = \frac{\partial}{\partial s} \int_{L(\theta,s)} f(x)\, \mathrm{d}x, \tag{1}$$

where $L(\theta, s) = \{x \in \mathbb{R}^2 : \ x_1 \cos\theta + x_2 \sin\theta = s\}$ denotes a line with the normal direction $(\cos\theta, \sin\theta)^T$ and the signed distance from the origin $s \in \mathbb{R}$. Given the measurements $y = \mathcal{P}f$, the reconstruction problem amounts to finding the refraction coefficient f from the phase shift data $\mathcal{P}f$. In this work, we consider the following noisy reconstruction problem

$$y^\delta = \mathcal{P}f + \eta, \tag{2}$$

where η is the noise component with a noise level $\delta > 0$, i.e., $\|\eta\|_2 \leq \delta$. To this end, we first note that the DPCI model (1) is essentially given by the first derivative of the classical Radon transform

$$\mathcal{R}f(\theta, s) = \int_{L(\theta,s)} f(x)\, \mathrm{d}x, \tag{3}$$

such that $\mathcal{P}f(\theta, s) = \frac{\partial}{\partial s}\mathcal{R}f(\theta, s)$. Therefore, techniques which were originally developed for conventional CT can be transferred to DPCI. For example, in [4], the well-known filtered backprojection (FBP) algorithms has been adapted for reconstruction from DPCI data. However, in order to achieve an adequate reconstruction quality, FBP needs a large number of projections. In addition to that, it is well-known that FBP performs poorly in the presence of noise.

To address these issues, we propose a reconstruction method that is based on a series expansion framework (often called algebraic or iterative reconstruction). In this framework, the unknown function f is expanded with respect to a given dictionary $(\psi_n)_{n=1}^N$ via $f = \sum_{n=1}^N c_n \psi_n$. Then, the expansion coefficients $(c_n)_{n=1}^N$ are determined from the measurements

$$y_m = \mathcal{P}f(\theta_m, s_m) = \sum_{n=1}^{N} c_n \mathcal{P}\psi_n(\theta_m, s_m). \tag{4}$$

by solving the linear system of equations

$$y = \mathbf{P}c, \tag{5}$$

where $y = (y_1, \ldots, y_M)^\intercal$ and $\mathbf{P} = (\mathbf{P}_{m,n})$ is the $M \times N$ *system matrix* with entries $\mathbf{P}_{m,n} = \mathcal{P}\psi_n(\theta_m, s_m)$.

Several choices of dictionaries have been proposed for reconstruction in DPCI. For example, in [5], Köhler et al. considered a series expansions based on Kaiser-Bessel functions (also known as isotropic blob functions). Another example is the B-Spline series expansion which was investigated by Nilchian et al. in [6].

In this work, we propose an approach that is based on the expansion of f with respect to the curvelet frame, [7]. An advantage of using curvelets over other dictionaries lies in the fact that curvelet expansions allow for an edge-preserving reconstruction. In addition to that, curvelets admit an analytic representation of the system matrix.

The paper is organized as follows: in Section II we recall the definition of curvelets and recall some of their basic properties. In Section III we describe our curvelet-based discretization used for the explicit computation of the system matrix $\mathbf{P}$ as well as the resulting reconstruction method. Finally, in Section

IV, we illustrate our method by reconstructions of simulated phantom data.

II. Curvelets

The curvelet dictionary is a family of functions $\psi_{j,l,k} : \mathbb{R}^2 \to \mathbb{C}$ which has a multi-scale structure and whose main advantage is the high directionality of its fine scale atoms [7]. The construction of curvelets is done in the Fourier domain. We will make use of the following definition of the Fourier transform

$$\hat{f}(\xi) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x) e^{-ix\xi} \, \mathrm{d}x.$$

We first define the generating curvelets $\psi_{j,0,0}$ at scale 2^{-j}, $j \in \mathbb{N}_0$, by using polar coordinates $\xi = re^{i\theta}$ in the Fourier domain:

$$\widehat{\psi}_{j,0,0}(re^{i\theta}) = 2^{-3j/4} \cdot W(2^{-j} \cdot r) \cdot V\left(\frac{2^{\lceil j/2 \rceil + 1}}{\pi} \cdot \theta\right), \quad (6)$$

where W is a radial window and V is an angular window, respectively. We require the windows W, V to be real and smooth ($W, V \in C^\infty$) such that $\operatorname{supp} W \subset (1/2, 2)$, $\operatorname{supp} V \subset (-1, 1)$. Moreover, V and W have to satisfy proper admissibility conditions, cf. [7]. The family of curvelets $\{\psi_{j,l,k}\}_{j,l,k}$ is constructed by translation and rotation of the generating curvelets $\psi_{j,0,0}$. That is, at scale 2^{-j}, the curvelet $\psi_{j,l,k}$ is defined via

$$\psi_{j,l,k}(x) = \psi_{j,0,0}(R_{\theta_{j,l}}(x - b_k^{j,l})), \text{ for } x \in \mathbb{R}^2. \quad (7)$$

Here $R_{\theta_{j,l}}$ denotes the rotation matrix

$$R_{\theta_{j,l}} = \begin{pmatrix} \cos\theta_{j,l} & -\sin\theta_{j,l} \\ \sin\theta_{j,l} & \cos\theta_{j,l} \end{pmatrix}$$

with respect to the scale-dependent rotation angles $\theta_{j,l}$ and the scale-dependent locations $b_k^{j,l}$ which are defined by

$$\theta_{j,l} = l \cdot \pi \cdot 2^{-\lceil j/2 \rceil - 1}, \quad -2^{\lceil j/2 \rceil + 1} \le l < 2^{\lceil j/2 \rceil + 1}, \quad (8)$$

$$b_k^{j,l} = R_{\theta_{j,l}}^{-1} \left(\frac{k_1}{2^j}, \frac{k_2}{2^{j/2}}\right), \quad k = (k_1, k_2) \in \mathbb{Z}^2. \quad (9)$$

Clearly, each curvelet is supported on a polar wedge in the Fourier domain which has a positive distance to the origin. We complete the curvelet system with the generating low-pass function $\psi_{-1,0,0}$, defined in the Fourier domain by

$$\widehat{\psi}_{-1,0,0}(re^{i\theta}) = W_0(r), \quad W_0^2(r) := 1 - \sum_{j=0}^{\infty} W^2(2^{-j} r),$$

with all of its translates $\{\psi_{-1,0,k}\}_{k \in \mathbb{Z}^2}$. The index set of the completed curvelet dictionary is now given by

$$\mathcal{I} = \{(-1, 0, k) : k \in \mathbb{Z}^2\} \cup \{(j, l, k) : j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2,\, -2^{\lceil j/2 \rceil + 1} \le l < 2^{\lceil j/2 \rceil + 1}\}$$

One of the fundamental properties of the curvelet dictionary is that it constitutes a normalized tight frame for $L^2(\mathbb{R}^2)$, cf. [7]. In particular, each $f \in L^2(\mathbb{R}^2)$ can be expanded in terms of curvelets via

$$f = \sum_{(j,l,k) \in \mathcal{I}} \langle \psi_{j,l,k}, f \rangle \, \psi_{j,l,k}. \quad (10)$$

Note that the representation (10) is directional. Indeed, in addition to the scale-parameter j and the location parameter $k = (k_1, k_2)$, the orientation parameter l corresponds to directional features of f.

III. Curvelet series expansion for DPCI

We now use curvelets in order to discretize the DPCI operator $\mathcal{P}$ which is defined in (1). To this end, we model f as a finite linear combination of curvelets via

$$f = \sum_{n=1}^{N} c_n \psi_n, \quad (11)$$

where $n = n(j, l, k)$ is an enumeration of the curvelet index set $\mathcal{I}$ and $N = |\mathcal{I}|$. In the following, we also assume that a finite number of measurements is available, $y_m = \mathcal{P}f(\theta_m, s_m)$, $1 \le m \le M \in \mathbb{N}$. Using (11), each measurement y_m can be expressed as

$$y_m = \mathcal{P}f(\theta_m, s_m) = \sum_{n=1}^{N} c_n \mathcal{P}\psi_n(\theta_m, s_m). \quad (12)$$

The discrete (noise free) reconstruction problem then reads

$$y = \mathbf{P}c, \quad (13)$$

where $\mathbf{P}$ is the system matrix which is defined by

$$\mathbf{P}_{m,n} = \mathcal{P}\psi_n(\theta_m, s_m), \quad 1 \le m \le M,\ n \in \mathcal{I}. \quad (14)$$

An advantage of using curvelets for the discretization lies in the fact that the Radon transform of curvelet elements can be computed analytically. In analogy to [8], we have the following result.

Theorem III.1. *Let $\psi_{j,l,k}$ be a curvelet (cf. (7)) and denote $e(\theta) = (\cos\theta, \sin\theta)^\intercal$. Then,*

$$\mathcal{P}\psi_{j,l,k}(\theta, s) = 2^{5j/4} V^*\left(\frac{2^{\lceil j/2 \rceil + 1}}{\pi}(\theta - \theta_{j,l})\right) \sqrt{2\pi} \, \widehat{iW^*}\left(2^j \left(\left\langle b_k^{j,l}, e(\theta - \theta_{j,l}) \right\rangle - s\right)\right)$$

where $b_k^{j,l}$ and $\theta_{j,l}$ are defined in (8) and (9). Further, $W^*(r) = rW(r)$ with the radial window function W as given above and $V^*(\alpha) = V(\alpha) + V(\alpha - \operatorname{sgn}(\alpha)\pi)$ with V denoting the angular window function.

We now consider the discrete noisy reconstruction problem

$$y^\delta = \mathbf{P}c + \eta, \quad (15)$$

where $\eta \in \mathbb{R}^N$ denotes the noise component. In order to minimize the influence of noise to the reconstruction, we use variational regularization which amounts to the minimization of an energy functional of the form

$$\|\mathbf{P}c - y\|_2^2 + \alpha\Lambda(c), \quad (16)$$

where the first term controls the data error and the second term $\Lambda : \mathbb{R}^N \to [0, \infty)$ is a penalty function which encodes the a-priori information about the unknown object f.

Our goal is to design an edge-preserving reconstruction method based on curvelet coefficients. For this purpose, we take advantage of the ability of curvelets to provide optimally sparse representation of functions with sharp edges, cf. [9]. In order to obtain a sparse vector of curvelet coefficients through minimization of (16), we use the paradigm of sparse regularization, cf.[10]. In this context, it is well-known that the ℓ^1-norm favors sparse solutions. Therefore, we solve the problem (15) by minimizing the ℓ^1-penalized functional

$$\hat{c} = \arg\min_{c \in \mathbb{C}^N} \left\{ \frac{1}{2} \left\| \mathbf{P}c - y^\delta \right\|_2^2 + \|c\|_{1,w} \right\}, \qquad (17)$$

where $\|c\|_{1,w} = \sum_k w_k |c_k|$ denotes the weighted 1-norm with a weight sequence w satisfying $w_k \geq w_0 > 0$. Having computed $\hat{c}$, a solution for the original problem (2) is then given by applying the synthesis operator to the regularized curvelet coefficients $\hat{c}$, i.e.,

$$f^* = \sum_{n=1}^{N} \hat{c}_n \psi_n. \qquad (18)$$

The computation of a reconstruction by (17), (18) is stable and edge-preserving.

IV. Results

A. Implementation

Our implementation of the curvelet transform is written in C++ and is based on polar coordinates in the Fourier domain. Note that the CurveLab toolbox [11] uses a slightly different approach with coronization based on concentric squares instead of concentric circles.

For sparse regularization, one of the most effective methods is the iterative soft-thresholding algorithm (ISTA) as suggested in [10]. We implemented the fast iterative soft-thresholding algorithm (FISTA) variant as proposed in [12]. In both algorithms every minimizer c^* of the ℓ_1-penalized reconstruction problem is a fixed point of the following iteration:

$$c^*_{k+1} = S_{\lambda w}\big(c^*_k - \lambda \mathbf{P}^*(\mathbf{P}c^*_k - y^\delta)\big).$$

Here, S_τ denotes the soft-thresholding operator of x with threshold τ:

$$S_\tau(x) = \begin{cases} x - \operatorname{sgn}(x)\tau & |x| \geq \tau \\ 0 & \text{else.} \end{cases}$$

The step-size $\lambda > 0$ is chosen according to the Barzilai-Borwein method (cf. [13]). For our first reconstructions, we use a constant parameter $w = 0.01$. We will refer to our ℓ_1-regularized reconstruction as Curvelet Sparse Regularization (CSR), cf. (17) and (18). Additionally, we implemented the Conjugate Gradient (CG) method. We have used the CG to solve the un-regularized tomographic problem (15). In the following, we will compare the results of reconstructions obtained via FBP to those of the CG based reconstruction as well as our CSR based reconstructions. In both implementations, CSR as well as CG, we have used the closed form formula given in Theorem III.1 in order to compute the elements of the system matrix.

B. Setup and datasets

To evaluate our method, we applied the forward model to the Shepp-Logan phantom. We compare FBP (with the adjusted filter for DPCI [4]) to both of our curvelet based iterative methods - CG and CSR. Both reconstructions, CG and CSR, were produced using 200 iterations. We start our evaluation by taking 360 projections within the angular range of 0° -180° . The results are shown in Fig. 1.

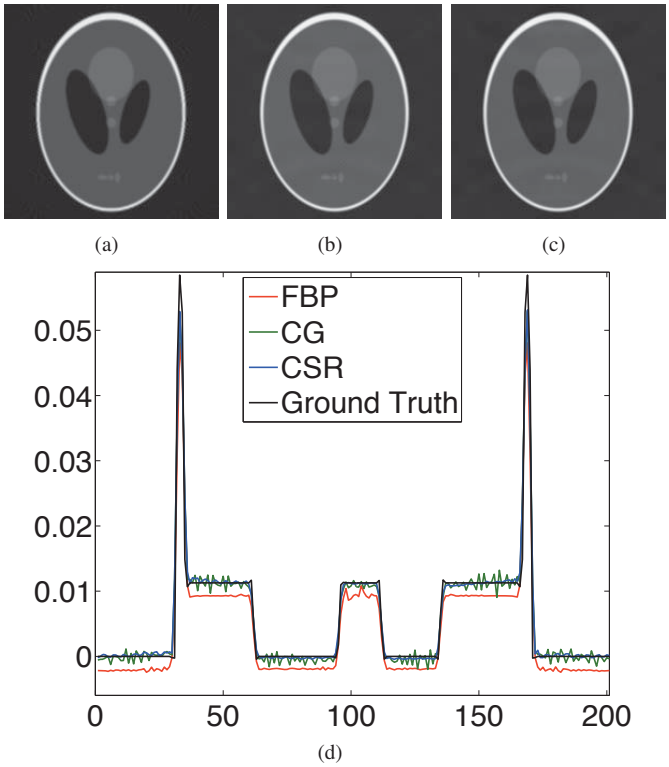


Fig. 1. (a) FBP (SNR: 12.7 dB), (b) CG (SNR: 15.7 dB), (c) CSR (SNR: 15.9 dB), (d) Line profile ($y = 100$)

The curvelet based methods provide visually promising results. Additionally, we present a line profile (see. Fig. 1 (d)) which clearly shows our curvelet based methods to be much closer to the ground truth than the FBP method. The signal to noise ratio (SNR) shows an increase from 12.7dB for FBP up to 15.9dB for CSR.

However, having a closer look we observe two additional phenomena: First, our current implementation of the curvelet based methods exhibits Gibbs-like effects. Second, we find the CG results to show high frequent noise in contrast to the FBP results. Thus, we suppose this noise is linked to the oscillating behaviour of the curvelet elements. This noise gets clearly reduced by the CSR method, while sharpness of the edges is preserved. The given line profile additionally shows, this noise reduction when using CSR in contrast to the CG method.

We continue our evaluation by reducing the number of projections taken to 60 by increasing the angular spacing from 0.5° to 3° . The corresponding results are shown in Fig. 2. The line profile in Fig. 2 (d) as well as the SNR increase from 11.1dB (FBP) to 14.3dB (CSR) emphasizes our previous results. However, in this case, we also observe Gibbs artefacts and a noise reduction when using the CSR method.

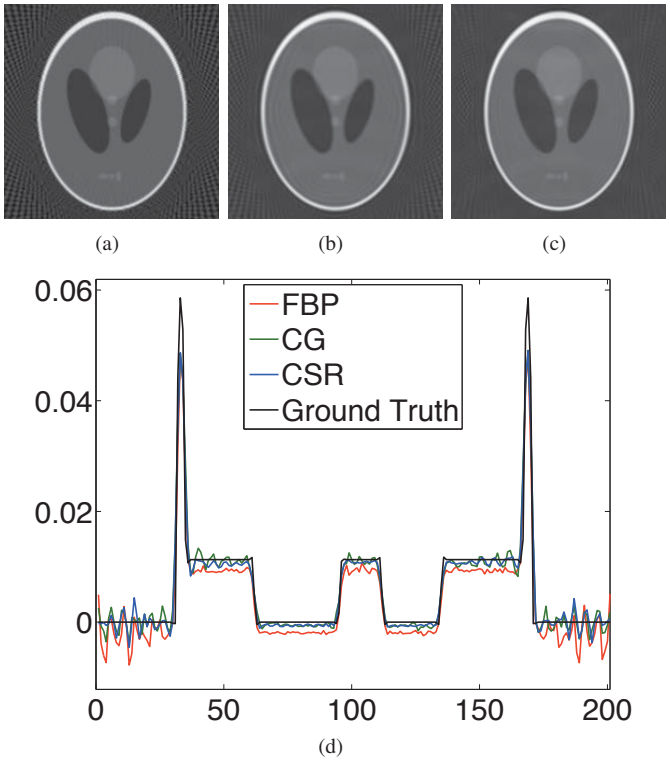


Fig. 2. (a) FBP (SNR: 11.1 dB), (b) CG (SNR: 13 dB), (c) CSR (SNR: 14.3 dB), (d) Line profile ($y = 100$)

V. Conclusion

In this work we have introduced a curvelet based framework for differential phase-contrast imaging. In particular, we have presented an analytically exact discretization of the DPCI operator. Moreover, we applied the curvelet sparse regularization to the reconstruction problem of DPCI. Our first reconstruction results show that the method is promising, but still requires further work to achieve better results.

We believe that the curvelet representation provides a solid basis for extending reconstruction methods towards reducing the radiation exposure by exploiting the sparsity and edge-preserving properties of curvelets.

Acknowledgment

This work was partially funded by the DFG Cluster of Excellence MAP and DFG SFB 824.

References

[1] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nature Physics*, vol. 2, no. 4, pp. 258–261, Mar. 2006.

[2] J. Kastner, B. Plank, and G. Requena, "Non-destructive characterisation of polymers and Al-alloys by polychromatic cone-beam phase contrast tomography," *Materials Characterization*, vol. 64, pp. 79–87, Feb. 2012.

[3] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. Brönnimann, C. Grünzweig, and C. David, "Hard-X-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, no. 2, pp. 134–137, Jan. 2008.

[4] F. Pfeiffer, O. Bunk, C. Kottler, and C. David, "Tomographic reconstruction of three-dimensional objects from hard X-ray differential phase contrast projection images," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 580, no. 2, pp. 925–928, Oct. 2007.

[5] T. Köhler, B. Brendel, and E. Roessl, "Iterative reconstruction for differential phase contrast imaging using spherically symmetric basis functions," *Medical Physics*, vol. 38, no. 8, p. 4542, 2011.

[6] M. Nilchian and M. Unser, "Differential phase-contrast X-ray computed tomography: From model discretization to image reconstruction," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE, 2012, pp. 90–93.

[7] E. J. Candès and D. L. Donoho, "Continuous curvelet transform. II. Discretization and Frames," *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 198–222, 2005.

[8] J. Frikel, "Sparse regularization in limited angle tomography," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 117–141, 2013.

[9] E. Candès and D. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2003.

[10] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[11] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast Discrete Curvelet Transforms," *Multiscale Modeling and Simulation*, vol. 5, no. 3, pp. 861–899, 2006.

[12] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, Jan. 2009.

[13] J. Barzilai and J. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 1988.

Directional X-Ray Scattering Tomography

Andreas Malecki, Guillaume Potdevin, Thomas Biernath, Elena Eggl, Tobias Lasser, Jens Maisenbacher, Jens Gibmeier, Alexander Wanner, and Franz Pfeiffer

***Abstract*—Since the introduction of dark-field imaging as a novel x-ray contrast method, a broad range of additional sample features have become accessible with laboratory-based tomography for the first time. Here contrast is caused by scattering inside the specimen and thus strongly depends on structures on a scale below the detector resolution. Simple x-ray dark-field tomography is in principle compatible to existing techniques with respect to the dose but only gives good reconstruction results when all structures inside the sample are isotropic meaning the scattering in each volume element does not change when recording the projections. Additionally a radiographic method exists that takes the changes in the scattering strength into account when rotating the sample: Directional dark-field radiography. But this method suffers from being applicable only to thin specimens. Here we present a novel method, directional x-ray scattering tomography, which combines directional dark-field imaging with a direction-aware reconstruction approach, and demonstrate its validity with experimental data from a well-defined specimen. With this novel method medical diagnosis based on non-resolvable structures for example in the case of bone strength could drastically improve.**

***Index Terms*—X-ray dark-field imaging, grating interferometry, scattering tensor tomography**

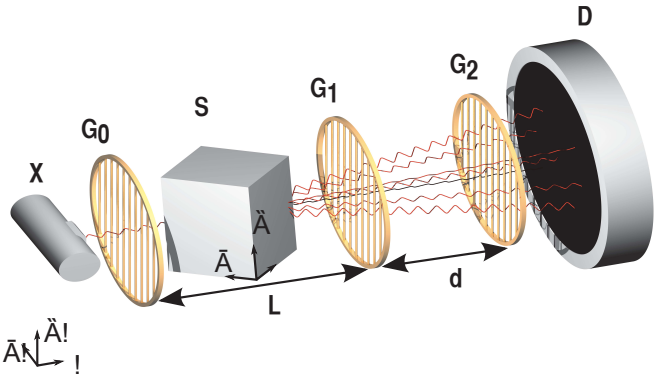


Fig. 1. **Sketch of the x-ray grating interferometry setup used to obtain the directional dark-field projections.** The setup consisted of an x-ray tube (X), a cubic specimen (S), a three-grating interferometer (G_0, G_1, and G_2) and a detector (D). The sample was rotated around a fixed point. It was mounted on an Eulerian cradle such that certain spatial angular positions had to be excluded for which the goniometer would shadow the cube partially or completely. The lab and sample coordinate systems are indicated as well.

I. Introduction

X-ray grating interferometry provides two novel contrast channels in addition to conventional absorption-based imaging: differential phase and dark-field contrast. In x-ray dark-field imaging [1] the signal represents the amount of coherent scattering of the penetrating x-ray wavefront at small structures inside the sample [2], [3]. Thus dark-field radiography and tomography allow to draw conclusions about the morphology of a sample's micro-structures [4], [5], [6], [7], [8], [9], [10], [11]. At the same time no high resolution detectors are required, which makes dark-field based methods especially interesting for medical applications because of almost equal dose requirements.

Whenever oriented sub-pixel size structures are present in the sample, the amount of scattering strongly depends on the current orientation of the specimen. Therefore the intensity of the dark-field signal oscillates when rotating of the specimen around the x-ray beam. Directional dark-field imaging [12], [13] evaluates these oscillations and allows to deduce the micro-structural orientation and anisotropy of thin specimens. Here it is crucial to examine only rather thin specimens where the x-ray beam then only penetrates one structure at maximum. Otherwise, whenever several structures overlap along the ray path and scatter in different directions, which is rather likely in thick specimens, further knowledge about the sample is required to distinguish between several contributing structures or a single one [14], [15]. The main limitation in this case is that it is unknown where along the beam path which contribution to the measured dark-field signal originates.

Directional dark-field imaging is a radiographic method. Computed tomography, however, allows to reconstruct position-dependent sample properties. As long as the scattering inside the sample is isotropic, tomographic reconstructions on the basis of the dark-field signal are feasible simply by replacing the measured transmission by the dark-field data [6]. But if the scattering inside the sample is anisotropic, neglecting the direction dependence of the scattering in dark-field tomography leads to incorrect reconstructions. When the sample is rotated around the tomographic axis, the amount of scattering strongly and non-linearly changes between subsequent projections. Consequently, coupling directional dark-field imaging to a direction-aware computed tomography method could overcome the limitations of both simple dark-field tomography and directional dark-field imaging.

Here we show how to combine the directional dark-field signal with computed tomography and reconstruct the local direction-dependent scattering of the sample. Based on experimental data obtained from a well-defined specimen we qualitatively account for the correct reconstruction of its

Andreas Malecki, Guillaume Potdevin, Thomas Biernath, Elena Eggl, and Franz Pfeiffer are with the Department of Physics and Institute of Medical Engineering, Technische Universität München, Boltzmannstr. 11, 85748 Garching, Germany.

Tobias Lasser is with the Chair for Computer Aided Medical Procedures (CAMP), Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany.

Jens Maisenbacher, Jens Gibmeier, and Alexander Wanner are with the Karlsruhe Institute of Technology, Institute for Applied Materials - WK, Engelbert-Arnold Strae 4, 76128 Karlsruhe, Germany.

Corresponding author: Andreas Malecki (Andreas.Malecki@ph.tum.de).

scattering parameters.

For this study we used a typical laboratory-based experimental setup (see fig. 1). It consists of an x-ray tube as source (X), a Talbot-Lau grating interferometer with three gratings (G_0, G_1, and G_2), and a detector (D). In such a setup, the first grating (source grating, G_0) ensures a sufficiently coherent illumination of the actual interferometer formed by the phase and analyzer gratings (G_1 and G_2). The sample (S) was mounted on an Eulerian cradle placed between source and phase grating. Consequently, its orientation could be chosen almost arbitrarily within a wide spatial angular range only limited by the goniometer's metal ring.

It has been shown before (cf. [14], [15]) that the physical model behind x-ray directional dark-field imaging is in a sense similar to the Beer-Lambert law known from absorption of x-rays and can be described by

$$D(x,y) = \exp\left[-\int \langle \boldsymbol{\epsilon}(x,y,z), \boldsymbol{t}\rangle^2 \,\mathrm{d}z\right]. \qquad (1)$$

Here D denotes the measured dark-field signal, and $\langle\cdot,\cdot\rangle$ the Euclidian scalar product, $\boldsymbol{\epsilon}(x,y,z) \in \mathbb{R}^3$ is the location-dependent scattering introduced by the sample, and $\boldsymbol{t} \in \mathbb{R}^3$ the sensitivity vector of the grating interferometer. $\boldsymbol{t}$ describes the absolute sensitivity and its direction at the same time. It always points in a direction parallel to the grating surface but perpendicular to the grating lines. The magnitude of the sensitivity vector depends on the setup parameters of the interferometer, the sample location, and the x-ray energy. The scattering as well is energy-dependent and characterized by its direction and magnitude represented by $\boldsymbol{\epsilon}$. For this work we ignored all deviations that may have been caused by the polychromatic energy spectrum and reconstructed effective values.

To combine x-ray directional dark-field imaging with computed tomography, we generalized the existing model in eq. (1) by allowing a finite number of different scattering directions $\boldsymbol{\epsilon}_i$ in every sample point. When performing tomography, the sample is rotated relative to the penetrating beam. As a consequence, the scattering at each location also depends on the direction of the incoming beam $\boldsymbol{s}$ relative to the sample orientation. These considerations lead to

$$D(x,y) = \exp\left[-\int \sum_i \langle \boldsymbol{\epsilon}_i(\boldsymbol{s},x,y,z), \boldsymbol{t}\rangle^2 \,\mathrm{d}z\right]. \qquad (2)$$

As already shown in [14], [15], it is impossible to distinguish between an isotropic scatterer and two completely anisotropic layers put close to each other but with a relative orientation of $90°$. Thus we can ignore constant scattering contributions by absorbing them in two perpendicular anisotropic contributions. We only have to ensure that enough different scattering directions are available.

To be able to perform a reconstruction, we still need to specify how $\boldsymbol{\epsilon}_i(\boldsymbol{s},x,y,z)$ changes when rotating the sample, which is equivalent to changing the direction of the incoming beam $\boldsymbol{s}$. Here we assume that the scattering direction is always fixed with respect to the sample coordinate system but the amount changes with the relative orientation of $\boldsymbol{\epsilon}_i$ and $\boldsymbol{s}$ like

$$\boldsymbol{\epsilon}_i(\boldsymbol{s},x,y,z) = |\hat{\boldsymbol{s}} \times \boldsymbol{\epsilon}_i(x,y,z)|\, \hat{\boldsymbol{\epsilon}}_i. \qquad (3)$$

Here $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{\epsilon}}_i$ are equal to $\boldsymbol{s}$ and $\boldsymbol{\epsilon}_i$ normalized to Euclidian unit length.

Using (3) we can substantiate (2):

$$D(x,y) = \exp\left[-\int \sum_i \left(|\boldsymbol{\epsilon}_i(x,y,z)|\,|\boldsymbol{t}|\right)^2 \left(|\hat{\boldsymbol{s}} \times \hat{\boldsymbol{\epsilon}}_i|\,\langle \hat{\boldsymbol{\epsilon}}_i, \hat{\boldsymbol{t}}\rangle\right)^2 \mathrm{d}z\right] \qquad (4)$$

All values of the weight factor $(|\hat{\boldsymbol{s}} \times \hat{\boldsymbol{\epsilon}}_i|\,\langle \hat{\boldsymbol{\epsilon}}_i, \hat{\boldsymbol{t}}\rangle)^2$ can be precalculated from the corresponding sample orientation when the setup geometry is known. For parallel geometry they are constant for all rays contributing to a single projection. The results presented below show the reconstruction of the product of the absolute values of $\boldsymbol{\epsilon}_i(x,y,z)$ and $\boldsymbol{t}$. This is equivalent to the assumption that all sample volume elements (voxels) contain an identical distribution of scatterers with respect to their size and material. If one can furthermore assume that the sensitivity is more or less constant and known, this directly allows to reconstruct the scattering caused in each sample voxel. In general this is not applicable, as the sensitivity strongly depends on the x-ray energy and structure dimensions. How to separate the two contributions from scattering and sensitivity and correctly model the energy dependence will be part of future examinations.

For the reconstruction we used the simultaneous algebraic reconstruction technique (SART) [16]. The underlying model was modified to fit to the experimental data by reconstructing several volumes in parallel, one for each scattering component (see Methods & Materials for details).

II. Results & Discussion

For this study we used a $1.03\,\mathrm{cm} \times 1.03\,\mathrm{cm} \times 1.02\,\mathrm{cm}$ large cube sample made of carbon fiber enforced carbon (CFC). The sample contained eight layers of oriented fibers with the fibers in each layer mainly pointing in one identical direction parallel to the layer. The layers were separated by thinner felt layers with an arbitrary fiber orientation. Groups of two equally oriented layers were oriented perpendicular towards each adjacent group and parallel to the upper/lower surface of the cube. From the top to the bottom all layers were connected by additional strings of fibers.

As the contribution of each scattering component depends on the relative orientation with respect to the beam direction, it was necessary to rotate the sample not only around a single rotation axis but freely around the center of rotation of the Eulerian cradle. Otherwise, there would have been scattering coefficients with a constant weight factor equal to 0. The orientation of the specimen in this case are defined by the three Eulerian angles θ, ψ, ϕ and in total 2416 different projections were recorded (see Methods & Materials for details).

Fig. 2 shows two representative slices of the raw results gained from reconstructing the scattering coefficients for the following seven different scattering orientations, along the three Cartesian coordinate axes and the four space diagonals (coordinates given with respect to the sample coordinate

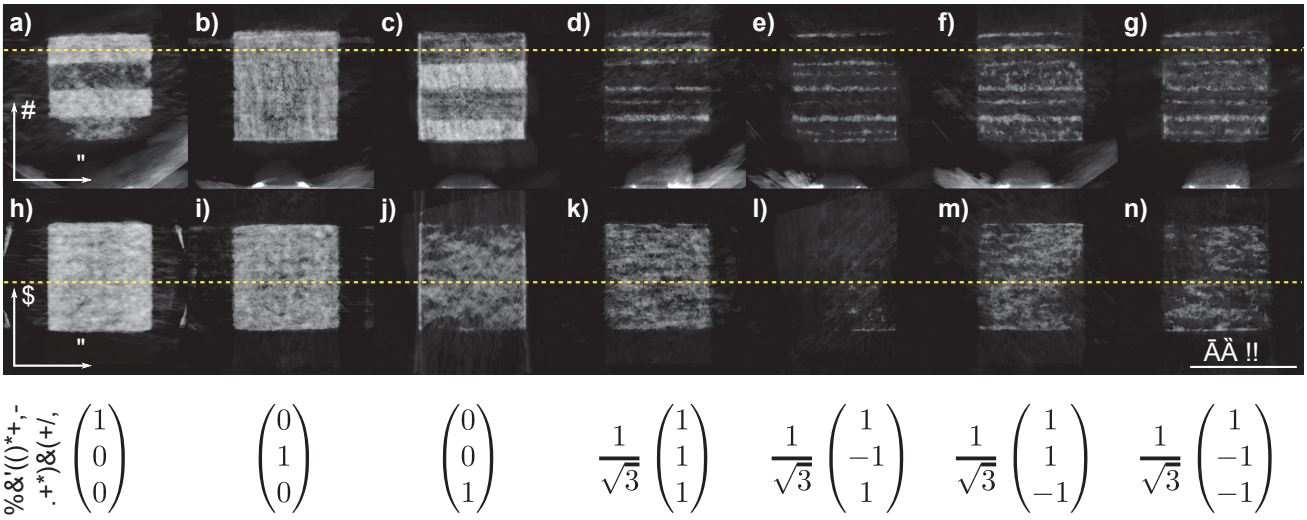


Fig. 2. **Reconstructed slices and direction-dependent scattering strength.** (a)-(g) show the scattering strength with respect to the scattering direction of the same central slice parallel to the sample's y-z plane. (h)-(n) show the same scattering components but now in a plane parallel to the x and z axes. The broken lines mark the position of the images shown. From top to bottom the strongest scattering direction inside the sample alternates between the x direction and the z direction. The scattering along the y axis is almost constant throughout the whole specimen. At the locations of the intermediate felt layers the the scattering power is equally distributed of over all scattering components.

system):

$$\hat{\boldsymbol{\epsilon}}_1 = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^T, \hat{\boldsymbol{\epsilon}}_2 = \begin{pmatrix}0 & 1 & 0\end{pmatrix}^T, \hat{\boldsymbol{\epsilon}}_3 = \begin{pmatrix}0 & 0 & 1\end{pmatrix}^T,$$
$$\hat{\boldsymbol{\epsilon}}_4 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\end{pmatrix}^T, \hat{\boldsymbol{\epsilon}}_5 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & -1 & 1\end{pmatrix}^T,$$
$$\hat{\boldsymbol{\epsilon}}_6 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & -1\end{pmatrix}^T, \hat{\boldsymbol{\epsilon}}_7 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & -1 & -1\end{pmatrix}^T.$$

We limited the number of scattering directions to seven to save memory and computation time. In the figure it becomes evident that the scattering happening inside the sample is mainly along the Cartesian coordinate axes and alternating between the x and the z direction, when moving along the y axis, while it is almost constant in the y direction throughout the whole specimen. Mainly only at the locations of the intermediate felt layers the diagonal scattering components show a significant contribution. For a fixed height y the scattering in all components is rather homogeneous.

In general scattering only occurs perpendicular to the structures in the sample. This leads to a strong dark-field signal pointing in these directions. From the reconstructed scattering components one can deduce that because of the maximum scattering strength alternating between the x and z direction, the main orientation of the structures between the corresponding layers must be alternating as well. Strictly speaking, the scattering structures in the uppermost layer and the second lowest have to point in the z direction and the features in the remaining two layers have to point in the x direction. As all fiber layers are stacked on top of each other in y direction, no change in the vertical scattering component can be observed. All these findings agree with the morphology of the sample resulting from its production process.

III. Conclusion & Outlook

In summary we have shown that the directional dark-field signal obtained with x-ray grating interferometry allows to tomographically reconstruct the local direction-dependent scattering properties of a specimen. For that purpose we extended the well-known physical model behind directional dark-field imaging to allow different scattering directions in every sample voxel.

In the future, this method can be extended to represent the scattering properties of the sample by a tensor in every sample voxel. If one takes the energy dependence of the scattering into account and energy-resolving pixelated detectors become available, it should also be possible to reconstruct the size distribution of the sub-pixel features causing the scattering. This technique in principle could also have a large impact in the medical field, for example in the diagnosis of bone pathologies like osteoporosis.

IV. Methods & Materials

The dark-field projections were obtained at a laboratory setup at the Technische Universität München (cf. sketch in fig. 1). We used a high power x-ray tube (MXR-160HP/11 by COMET AG, Switzerland) at $90\,\mathrm{kV}$ and $8\,\mathrm{mA}$ and an additional $3.0\,\mathrm{mm}$ aluminium filter. The x-ray detector was a Varian PaxScan 2520D with a CsI scintillator and a pixel pitch of $127\,\mu\mathrm{m}$ (Varian medical systems, USA).

The interferometer consisted of two transmission gratings (G_0 and G_2) with a silicon substrate height of $500\,\mu\mathrm{m}$ and $150\,\mu\mathrm{m}$ and $160 - 170\,\mu\mathrm{m}$ high gold lines and spaces filled with SU-8. Our phase-shifting grating (G_1) induced a phase shift of $\pi/2$ at the targeted design energy of $56.9\,\mathrm{keV}$ and was made of $10\,\mu\mathrm{m}$ high nickel lines on a $200\,\mu\mathrm{m}$ thick silicon substrate. We used a symmetric setup with an inter-grating distance of $L = d = 1.15\,\mathrm{m}$. The period of the gratings was $10\,\mu\mathrm{m}$ for the absorption gratings and $5\,\mu\mathrm{m}$ for the phase grating and all duty cycles were 0.5.

The sample was mounted on an Eulerian cradle manufactured by Huber Diffraktionstechnik GmbH & Co. KG,

Germany, and located $89\,\mathrm{cm}$ from the source grating. With a resulting magnification of 2.6 and a sample size of $1\,\mathrm{cm}$ this allowed to use parallel beam geometry as a very good approximation for this reconstruction problem.

Projections were taken for the following Eulerian angle values given as (θ, ψ, ϕ) with coarse steps in θ and ϕ and fine steps in ψ. With x, y, z being the Cartesian sample coordinate axes and x', y', z' the Cartesian lab coordinate system, θ was the angle between the y' and y axis, ψ the rotation angle around y' and ϕ the rotation angle around y. During the experiment ψ was used as fastest and ϕ as slowest axis. This resulted in a total number of 2416 projections. For each sample orientation the phase grating was scanned in 8 steps over one grating period (phase stepping, cf. [17]). Reference images were taken each time between a full scan along the fast axis. We calculated the dark-field contrast signal from the Fourier transform of the resulting intensity variation (cf. [17], [1]).

The carbon cube specimen was created from several layers of carbon fibers, which were then infiltrated with the carbon matrix by chemical vapor infiltration (CVI). This was done at the Institute for Chemical Technology and Polymer Chemistry at the Karlsruhe Institute of Technology (KIT).

Blocks of $15 \times 50 \times 55\,\mathrm{mm}^3$ size were streamed under isothermal and isobaric conditions with methane. The specific correlations between infiltration parameters and resulting carbon-carbon structures can be found in [18]. The specimen researched in this work has been infiltrated at $1095\,^\circ\mathrm{C}$ and $25\,\mathrm{kPa}$ for $90\,\mathrm{h}$ with a dwell time of $0.1\,\mathrm{s}$. The carbon fiber laminate is commercially available ("Surface Transforms", Ellesmare Port, UK) and consists of HT-fibres (Panox®, SGL Carbon) with a Young's modulus of $190\,\mathrm{GPa}$ and a density of $1.72\,\mathrm{g/cm}^3$. From one of the large carbon blocks a cube of size $1.03\,\mathrm{cm} \times 1.03\,\mathrm{cm} \times 1.02\,\mathrm{cm}$ was cut, which served as specimen in our study. It contained 4 groups of 2 adjacent layers with alternating orientation between the groups as described above.

As already mentioned above we used the SART algorithm for reconstruction. We stopped after 4 iterations, after which no further significant changes were introduced to the result. For forward and backward projection we used trilinear interpolation. The reconstructed volume contained $358 \times 358 \times 358$ volume elements with an edge length of $49\,\mu\mathrm{m}$ each.

Acknowledgments

We acknowledge fruitful discussions with Andreas Fehringer, Dieter Hahn, Martin Bech and Pierre Thibault, and financial support through the DFG Cluster of Excellence Munich-Centre for Advanced Photonics (MAP), the DFG Gottfried Wilhelm Leibniz program, the DFG grant BA 4085/1-2, and the European Research Council (ERC, FP7, StG 240142). This work was carried out with the support of the Karlsruhe Nano Micro Facility (KNMF, www.kit.edu/knmf), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT). A.M. acknowledges the TUM Graduate School for the support of his studies. A.M., T.B., J.M., and J.G. appreciate the support of the virtual institute "New X-ray analytic methods in material science" of the Helmholtz association.

References

[1] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. Brönnimann, C. Grünzweig, and C. David, "Hard-x-ray dark-field imaging using a grating interferometer," *Nat. Mater.*, vol. 7, no. 2, pp. 134–137, 2008.

[2] W. Yashiro, Y. Terui, K. Kawabata, and A. Momose, "On the origin of visibility contrast in x-ray talbot interferometry," *Opt. Express*, vol. 18, no. 16, pp. 16 890–16 901, 2010.

[3] S. K. Lynch, V. Pai, J. Auxier, A. F. Stein, E. E. Bennett, C. K. Kemble, X. Xiao, W.-K. Lee, N. Y. Morgan, and H. Harold Wen, "Interpretation of dark-field contrast and particle-size selectivity in grating interferometers," *Appl. Opt.*, vol. 50, no. 22, pp. 4310–4319, Jul. 2011.

[4] H. Wen, E. E. Bennett, M. M. Hegedus, and S. Rapacchi, "Fourier x-ray scattering radiography yields bone structural information." *Radiology*, vol. 251, no. 3, pp. 910–918, Jun. 2009.

[5] G. Potdevin, A. Malecki, T. Biernath, M. Bech, T. H. Jensen, R. Feidenhans'l, I. Zanette, T. Weitkamp, J. Kenntner, J. Mohr, P. Roschger, M. Kerschnitzki, W. Wagermaier, K. Klaushofer, P. Fratzl, and F. Pfeiffer, "X-ray vector radiography for bone micro-architecture diagnostics." *Phys. Med. Biol.*, vol. 57, no. 11, pp. 3451–3461, jun 2012.

[6] M. Bech, O. Bunk, T. Donath, R. Feidenhans'l, C. David, and F. Pfeiffer, "Quantitative x-ray dark-field computed tomography," *Phys. Med. Biol.*, vol. 55, no. 18, pp. 5529–5539, Sep. 2010.

[7] G.-H. Chen, N. Bevins, J. Zambelli, and Z. Qi, "Small-angle scattering computed tomography (sas-ct) using a talbot-lau interferometer and a rotating anode x-ray tube: theory and experiments." *Opt. Express*, vol. 18, no. 12, pp. 12 960–12 970, Jun. 2010.

[8] W. Yashiro, S. Harasse, K. Kawabata, H. Kuwabara, T. Yamazaki, and A. Momose, "Distribution of unresolvable anisotropic microstructures revealed in visibility-contrast images using x-ray talbot interferometry," *Phys. Rev. B*, vol. 84, no. 9, p. 094106, Sep. 2011.

[9] M. Chabior, T. Donath, C. David, M. Schuster, C. Schroer, and F. Pfeiffer, "Signal-to-noise ratio in x ray dark-field imaging using a grating interferometer," *J. Appl. Phys.*, vol. 110, no. 5, p. 053105, 2011.

[10] P. Modregger, F. Scattarella, B. Pinzer, C. David, R. Bellotti, and M. Stampanoni, "Imaging the ultrasmall-angle x-ray scattering distribution with grating interferometry," *Phys. Rev. Lett.*, vol. 108, no. 4, p. 048101, Jan. 2012.

[11] A. Malecki, G. Potdevin, and F. Pfeiffer, "Quantitative wave-optical numerical analysis of the dark-field signal in grating-based x-ray interferometry," *Europhys. Lett.*, vol. 99, no. 4, p. 48001, aug 2012.

[12] T. H. Jensen, M. Bech, O. Bunk, T. Donath, C. David, R. Feidenhans'l, and F. Pfeiffer, "Directional x-ray dark-field imaging," *Phys. Med. Biol.*, vol. 55, no. 12, pp. 3317–3323, May 2010.

[13] T. H. Jensen, M. Bech, I. Zanette, T. Weitkamp, C. David, H. Deyhle, S. Rutishauser, E. Reznikova, J. Mohr, R. Feidenhans'l, and F. Pfeiffer, "Directional x-ray dark-field imaging of strongly ordered systems," *Phys. Rev. B*, vol. 82, no. 21, p. 214103, Dec. 2010.

[14] V. Revol, C. Kottler, R. Kaufmann, A. Neels, and A. Dommann, "Orientation-selective X-ray dark field imaging of ordered systems," *Journal of Applied Physics*, vol. 112, no. 11, p. 114903, 2012.

[15] A. Malecki, G. Potdevin, T. Biernath, E. Eggl, E. Grande Garcia, T. Baum, P. B. Noël, J. S. Bauer, and F. Pfeiffer, "Coherent superposition in grating-based directional dark-field imaging," submitted to PLOS ONE.

[16] A. H. Andersen and A. C. Kak, "SIMULTANEOUS ALGEBRAIC RECONSTRUCTION TECHNIQUE (SART): A SUPERIOR IMPLEMENTATION OF THE ART ALGORITHM," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[17] T. Weitkamp, A. Diaz, C. David, F. Pfeiffer, M. Stampanoni, P. Cloetens, and E. Ziegler, "X-ray phase imaging with a grating interferometer," *Opt. Express*, vol. 13, no. 16, p. 6296, 2005.

[18] W. G. Zhang, Z. J. Hu, and K. J. Hüttinger, "Chemical vapor infiltration of carbon fiber felt: optimization of densification and carbon microstructure," *Carbon*, vol. 40, pp. 2529–2545, 2002.

Reducing Metal Artifacts Based on Three Approximately Orthogonal Projections: Preliminary Results in Dental CBCT with a Half-size Detector

Qingli Wang, Liang Li, Li Zhang*

***Abstract*—Metallic implants will cause serious artifacts in CT images. The artifacts may increase the difficulty of diagnosis, especially in dental CBCT where metallic implants appear in more than 50% clinical cases. A novel metal artifact reducing method based on three approximately orthogonal projections has been presented and proved to be effective by experimental data. Here in this article, we improve this method to adapt to the dental CBCT with a half-size detector. Scanning data from clinical case was used to illustrate the effect of the method.**

***Index Terms*—Dental CBCT, half-size detector, metal artifact.**

I. Introduction

METAL artifacts may seriously degrade the quality of reconstruction images in medical CT system, increasing the difficulty of diagnosis. Especially in dental CT, over 50% of the clinical cases are with metallic implants. The shape and the material of the metallic implants are multifarious, and the spatial distribution is always complicated. These may increase the difficulty of algorithm design.

Several metal artifact reducing methods have been published in past years, which can be classified into three groups: projection data pre-processing methods[1][2][3][4], iterative reconstruction methods[5][6] and image post-processing methods[7]. In 1987, Kalender et al created a precedent for reducing metal artifacts in 2D CT images, using the projection data pre-processing method[1]. They first located the metal part in the projection data, and then modified the metal projection data using the interpolation algorithm. The CT image was reconstructed from the modified data, and the metal part would be pasted back. This idea has been developed to deal with the metal artifacts in different types of CT systems ever since. However, most of these methods require pre-reconstruction, which would reduce the computing speed, lowering the practicality of the methods in CBCT system.

This work is supported by National Key Technology R&D Program of the Ministry of Science and Technology (No. 2012BAI07B05)

All the authors are with 1) Department of Engineering Physics, Tsinghua University, Beijing, 100084. China. 2) Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing, 100084. China

* Li Zhang, Email: zli@tsinghua.edu.cn.

In 2007, Zhang et al proposed a method in which the user needs to segment the metal areas manually in the projection images at two chosen angles[3]. The method located the metallic implants in the 3D space by geometry calculation, and then segmented and modified the metal area in each projection image using the metal locations as a priori information. This method has been proved to be as efficient as the pre-reconstruction methods in the clinical experiment, but with lower computational cost. However, the mode of human-computer interaction means the user's experience may affect the result a lot.

A metal artifact reducing method based on three approximately orthogonal images was presented in 2012[4]. This method first reconstructs the parallel-beam projection image along the Z-axial direction of the FOV. This parallel-beam projection image will help to choose the two projection images in which the metal areas do not overlap. Then the two chosen projection images and the overhead view image will be used together to calculate the 3D coordinates of the metallic implants. The classical region growing method and the bilinear interpolation method are used to modify the projection data of the metal areas in each projection image. This method can run automatically and has been proved to be efficient by experimental data of a simple model.

A CBCT system with a half-size detector has been developed to be applied in clinical dental diagnosis. As shown in Fig. 1, the application of the half-size detector in CBCT can expand the imaging field of view (FOV), and reduce the manufacturing cost of the system. Here in this paper, we improve the method mentioned in the previous paragraph in some key steps to adapt to the application of the dental CBCT with a half-size detector. Scanning data from clinical case with metallic implants was used to illustrate the effect of the method.

II. Material and Methods

A. Systems

Fig. 1 shows the geometry of the CBCT system with a half-size detector. $Oxyz$ represents a Cartesian coordinate system. $S(\beta)$ is the focal spot of the X-ray source, where β is the angular parameter. The flat panel detector is a $100mm \times 100mm$ square with 512×512 pixels. The straight line

connecting the X-ray source and the center of rotation passes through the pixel at Row 256, Column 499 of the detector vertically. Previous researches show that the half-detector CBCT system can accurately reconstruct examined objects using data of 360-degree projections[8].

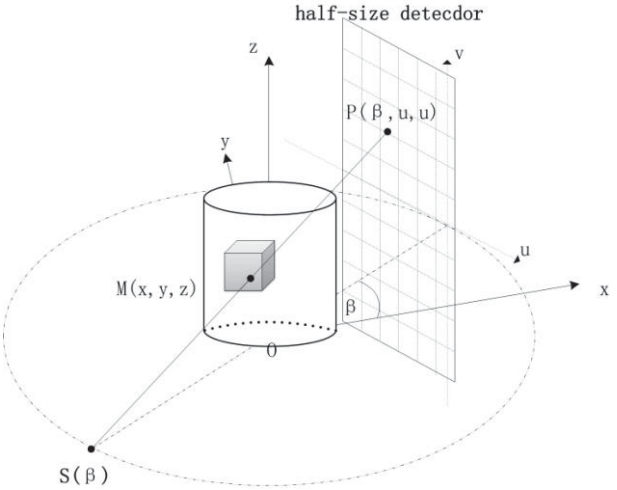


Fig. 1. A schematic diagram showing the half-size detector CBCT system.

B. Metal Artifact Reducing Method Based on Three Approximately Orthogonal Projections

A metal artifact reducing method based on three approximately orthogonal projections has been presented, providing a fast solution to reducing the metal artifact in the CBCT system[4]. This method can be achieved in practical applications through the following four steps.

Step 1: Synthesis the parallel-beam projection image from the overhead view.

Li et al proposed an exact method to synthesize parallel-beam projections along the Z-axis from cone-beam projections[9]. The algorithm includes two steps. First, the derivatives of Radon data in the plane perpendicular to the Z-axis can be computed according to equation (1),

$$\frac{\partial}{\partial u} R_f\left(s,\vec{m}\right)=\frac{1}{\cos^2\gamma}\int_{t_1}^{t_2}\frac{\partial}{\partial u}\left(\frac{SO}{SA(t)}\times P(\beta,u,v)\right)dt \qquad (1)$$

As it is shown in Figure 2, C is a point within the horizontal surface Ω, and s represents the length of OC, while $\vec{m}$ represents the unit victor of $\overrightarrow{OC}$. $R_f\left(s,\vec{m}\right)$ is the corresponding Radon value of the parallel-beam projection along the Z-axis. $P(\beta,u,v)$ is the projection data of the ray SA.

Then, we may use a BPF method (as shown in equation (2)) to synthesize the parallel-beam projection image along the Z-axis,

$$Pf(\vec{r})=\frac{1}{2\pi}\int_0^{\pi}\left[\frac{\partial Rf(s)}{\partial s}*h_H(s)\right]\Bigg|_{s=\vec{r}\cdot\vec{m}} d\beta \qquad (2)$$

where * denotes the convolution operation and $h_H(s)$ is the Hilbert filter defined as

$$h_H(s)=\int_{-\infty}^{+\infty}\left(-\mathrm{i}\cdot\mathrm{sign}(\rho)\right)\cdot e^{2\pi i\rho s}\mathrm{d}\rho \qquad (3)$$

This step can be transplanted to the half-size detector CBCT system. In the clinical applications, some improvement is necessary, which will be introduced hereinafter.

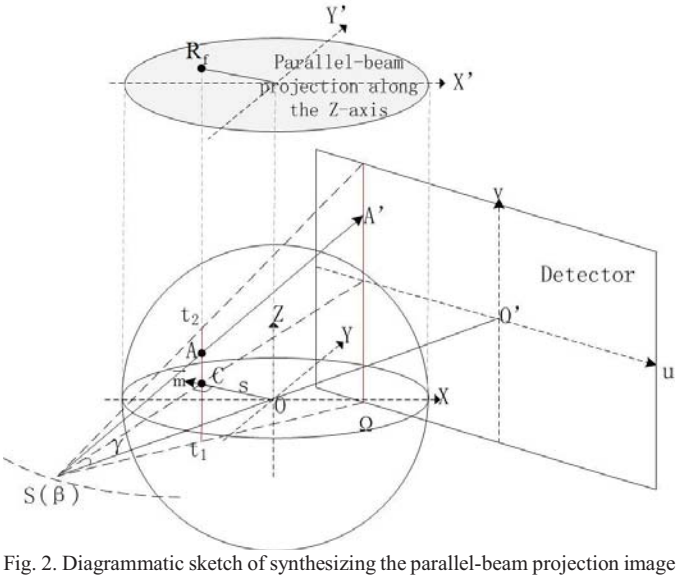


Fig. 2. Diagrammatic sketch of synthesizing the parallel-beam projection image from the overhead view in circular CBCT.

Step 2: Locate the metallic implants in the FOV.

In this method, a point inside the metallic implant named 'metal anchor point' (MAP), is used to locate the metallic implant in the 3D space. To locate the MAPs in the FOV, the metal areas are first marked in the overhead view image by a thresholding segmentation method. Then the overhead projection image is used to choose the other two horizontal cone-beam projection images from the scanned data in which the projection areas of different metallic implants are not overlapping, or the overlapping areas are as small as possible. These three projection images (one overhead image and two horizontal cone-beam projection images) will be used to calculate the 3D coordinates of the MAPs according to the CBCT scanning geometry. The detail of this MAR locating method can be found in the reference[4].

Step 3: Segment and modify the metal areas in projection images.

Once the 3D coordinates of all the MAPs are obtained, the forward projections of the MAPs, which named metal seed points (MSPs), can be located in each projection image. Since these MSPs are necessarily located within the metal projection areas, the classical region growing method[10] can be executed to segment the metal areas out of the projection image, using the MSPs as a priori information.

After marking all the metal areas out of each projection image, a bilinear interpolation algorithm is executed to modify the projection data of the metal areas.

This step should be modified in CBCT system with half-size detector, which will be introduced later on.

Step 4: Reconstruct the CT image.

The FDK algorithm is used to reconstruct the CT image from the above modified projection data. This method can also be used in the CBCT system with a half-size detector.

C. Synthesis the overhead view image

The method mentioned above, which synthesis the overhead view image, can be implemented in the CBCT system with a half-size detector.

If the projection data are truncated along the Z-axis, the overhead view image cannot be mathematically exact. However, the image which is approximately calculated can also be used for locating metallic implants, as long as all the metallic implants are in the FOV and the contrast of the metallic implants in the overhead view image is high enough to be identified.

In the clinical applications, patient's head is fixed on the carriage, so the metallic implants on the teeth are always in a certain height range. In equation (1), we can select the value of t_1 and t_2, to make sure that all the metallic implants has been contained in the overhead view image. Meanwhile, as shown in Fig. 3, the line t_1t_2 should be as short as possible, to enhance the contrast of the metallic implants in the overhead view image.

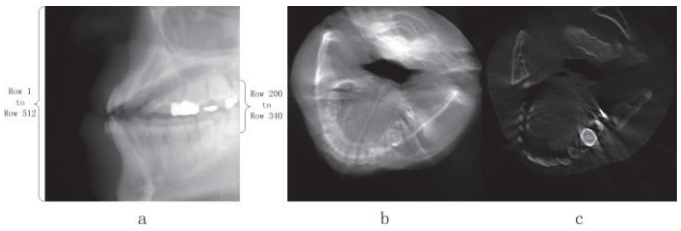


Fig. 3. (a) is a projection image. (b) is the overhead view image synthesized by the data from Row 1 to Row 512, while (c) is the overhead view image synthesized by the data from Row 200 to Row 340. The metals in image (c) are more prominent.

D. Locate the MSP

The MSPs can be located by the forward projections of the MAPs. Then the region growing method and the bilinear interpolation can be used to modify the metal areas of the projection images. However, as shown in Fig. 1, all the projection images are truncated at the right side of the half-size detector, so metallic implants would be truncated inevitably in some of the projection images.

Sometimes when the metallic implant is truncated, the MSP may be projected outside the detector (as the point A shown in Fig. 3 (c)). A new seed point is necessary to support the region growing method, which would be located by the following steps:

Step 1: If the distance between A and the detector boundary is less than 30 pixels, we first locate the pixel B in the same row with A at the boundary of the detector.

Step 2: Locate the pixel corresponding the maximum gray value from the region with red shadow, and name the pixel as C. The pixels of the red shadow region are all at the boundary, and the distances between these pixels and B are less than 5 pixels.

Step 3: C would be regarded as the new seed point if the gray value of C is higher than a pre-selected threshold. Otherwise, there is no metallic implant truncated in this region. In this work, the threshold is set at 4 empirically.

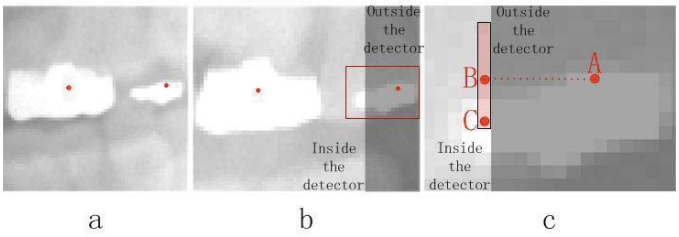


Fig. 4. Both of the MSPs are projected in the metal areas in image (a), while one of the MAPs is outside the detector in image (b). (c) is the partial enlarged view of the region where metal is truncated.

III. EXPERIMENT AND RESULTS

The experimental data was collected from our half-size detector dental CBCT system. The flat panel detector is a $100mm \times 100mm$ square with 512×512 pixels. The straight line connecting the X-ray source and the center of rotation passes through the pixel at Row 256, Column 499 of the detector vertically. The distance between the X-ray source and the rotation axis $R = 700mm$, while the distance between the X-ray source and the detector $D = 1000mm$. The voltage and the current of the X-ray tube are $100keV$ and $4mA$ respectively. As the system rotates, the detector collects data at intervals of one degree, so the whole data to be processed include 360 projection images. The scanning data are from clinical cases of volunteers.

We first synthesize the parallel-beam projection image from the overhead view, using the projection data between Row 200 and Row 340 on the detector. A thresholding segmentation method is used to mark the metal areas in the overhead view image. Two pairs of conjugated horizontal projection images, in which the metallic implants are not overlapping, are chosen from the certain angles determined by the segmented overhead view image. As shown in Fig. 5, the five segmented projection images would be used to locate the MAPs in the 3D space. The coordinates of the MAPs are shown in TABLE I.

TABLE I
COORDINATES OF THE MAPS

Num.	3-D coordinates (x,y,z) / (mm)
1	(5.92, -22.67, -1.81)
2	(13.59, -15.06, -1.82)
3	(19.49, -7.85, 0.48)

With the help of the MAPs, we can locate the MSPs in each cone-beam projection image. The region growing segmentation method can mark the metal areas base on the locations of the MSPs, and the bilinear interpolation method can modified the projection data of the metal areas, as shown in Fig. 6.

Results obtained by our metal artifacts reducing method are shown in Fig. 7. The streak artifacts and the cupping artifacts are suppressed. But some original structure has been blurred.

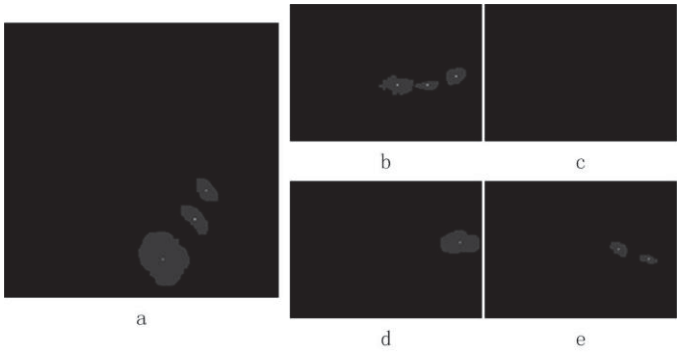


Fig. 5. (a) is the segmented overhead view image, while (b), (c), (d), (e) are the two pairs of conjugated horizontal projection images, from the projection angle $\beta_1 = 5^\circ$, $\beta_2 = 185^\circ$, $\beta_3 = 65^\circ$ and $\beta_4 = 245^\circ$. The mass center of each metal area is marked as highlight in the images.

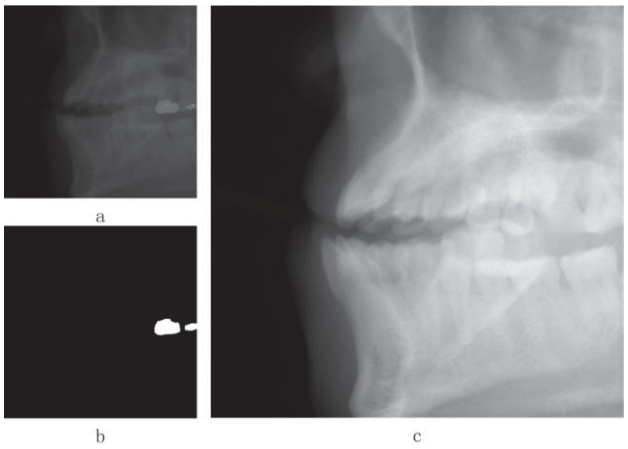


Fig. 6. (a) shows the MSPs as highlight in the cone-beam projection image. (b) is the result of the region growing segmentation method. (c) is the modified result of the bilinear interpolation method.

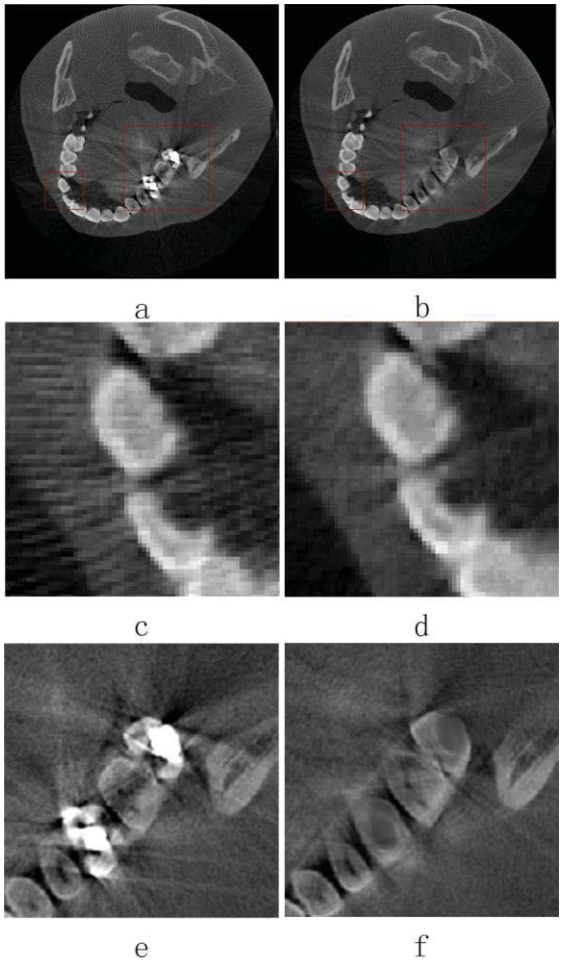


Fig. 7. (a) is reconstructed from the original projection data, while (b) is reconstructed from the modified data. (c), (d), (e) and (f) are the corresponding partial enlarged view images.

IV. Discussion and Conclusion

The metal artifact reducing method based on three approximately orthogonal images provides a fast solution to reduce the metal artifacts in the CBCT system without pre-reconstruction. Here we improve the method to adapt to the clinical applications in the dental CBCT with a half-size detector. The experimental results of the scanning data from the clinical cases show that this method can suppress the streak artifacts and the cupping artifacts. However, because of the error introduced by the segmentation and interpolation algorithm, some original structures may be blurred synchronously.

In order to make this method more practical, some further researches need to be focus on the following issues. First, in the clinical applications of the dental CBCT, it is difficult for the region growing algorithm to ensure a high accuracy. Better segmentation algorithm is the key to improve the performance of the method. Furthermore, bilinear interpolation algorithm cannot ensure the image is smooth at the boundary of the interpolating region, and this may introduce radial artifacts in the reconstruction images, blurring the original structures. Some more practical interpolation methods should be introduced to make the modification of the projection images closer to reality.

References

[1] W. A. Kalender et al, Reduction of CT artifacts caused by metallic implants, Radiology, vol. 164, pp. 576-577, Aug. 1987.

[2] Meilinger M. et al, Metal artifact reduction in cone beam computed tomography using forward projected reconstruction information, Z Med. Phys. vol. 21, pp. 174-82. Sep. 2011.

[3] Yongbin Zhang et al, Reducing metal artifacts in cone-beam CT images by preprocessing projection data, Radiation Oncology Biol. Phys., Vol. 67, pp. 924–932, 2007.

[4] Qingli Wang et al, A novel metal artifact reducing method for cone-beam CT based on three approximately orthogonal projections, Phys. Med. Biol. Vol. 58, pp. 1–17, Jan. 2013

[5] Ge Wang et al, Iterative deblurring for metal artifact reduction, IEEE Trans. Med. Imaging, vol. 15, pp. 657-667, 1996.

[6] B. De Man et al, An iterative maximum-likelihood polychromatic algorithm for CT, IEEE Trans. Med. Imaging, vol. 20, pp. 999-1008, 2001.

[7] V. Naranjo et al, A new 3D paradigm for metal artifact reduction in dental CT, IEEE International Conference on Image Processing (ICIP), (Brussels, Belgium, 11-14, Sep. 2011), pp. 461-464.

[8] Liang L, et al. A New Cone-beam X-ray CT System with a Reduced Size Planar Detector. Chinese Physics Society C, vol. 30, pp. 812-817,2006

[9] Liang Li et al, A general exact method for synthesizing parallel-beam projections from cone-beam projections via filtered back-projection Phys. Med. Biol. vol. 51, pp. 5643-5654, Oct 2006.

[10] R. Adams et al, Seeded region growing, IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, pp. 641-647, Jun. 1994.

Tomographic Image Reconstruction on Mixed Geometry Finite Element Meshes

Rostyslav Boutchko[1], Alexander I. Veress, and Grant T. Gullberg

***Abstract*—We present an approach that combines the application of ECG-gated single photon emission tomographic imaging of the heart with finite-element mesh modeling of the heart. ECG-gated image datasets of the heart have been used to create a hexahedron-based FE model of the left ventricle. This mesh has been previously used to model the work of the heart muscle. We propose to combine the validated moving hexahedral mesh occupying the space of the heart muscle with a deformable tetrahedral mesh occupying the field of view of the scanner detector. The resulting mesh is used as an image representation for the radiotracer concentration distribution function. The tomographic problem for the constant intensities inside each polyhedron of the mesh is solved using ML-EM algorithm. Since the underlying hexahedral mesh is designed to model the cardiac motion, the new image representation allows us to reconstruct a single intensity value per volume element using projection data from all cardiac gates, thus improving the reconstructed image quality. After the motion-corrected 3D image is computed, it is used to re-assess the input data and improve both the mesh geometry and the cardiac motion modeling. Repeating this process iteratively, we aim to determine the best mesh-based representation and the most accurate motion model from the initially acquired projection dataset. This approach is directed toward enabling automatic generation of the initial FE model of the heart. The main challenges encountered in the work are achieving a successful combination of a hexahedral and a tetrahedral mesh and in preserving the topological integrity of the mesh during motion.**

***Index Terms*—SPECT, cardiac imaging, finite element modeling, non-uniform meshes, multiresolution, image reconstruction, motion corrected image reconstruction.**

I. Introduction

CARDIAC single photon emission computed tomography (SPECT) is a method of determining three-dimensional distributions of injected radiotracers with selected biochemical properties inside the patient's heart. Wide access to SPECT scanners and availability of a variety of radiotracers makes SPECT one of the most popular cardiac diagnostic modalities. The main challenge of cardiac SPECT (as well as another nuclear emission imaging modality, PET) is the low quality of the reconstructed images manifested in terms of image artifacts. The major sources of imaging artifacts are high noise content in the acquired data and patient motion. Reducing image artifacts caused by these factors is one of the primary goals of modern nuclear emission technology research.

The noise content in the data acquired in emission tomography is determined by the dose of the injected radiotracer, camera sensitivity, and the imaging time, thus the noisiness of the reconstructed images has to be reduced by data processing tools. A promising approach to reducing noise content in the target images is using image representation methods that permit spatially varying resolution, e.g. larger voxel sizes in those part of the image where the spatial resolution can be partially compromised. Triangular (in 2D) and tetrahedral (in 3D) meshes of non-uniform density have been used as multiresolution image representation in the past [1-6]. These meshes offer both a multiresolution data representations and a convenient tool for incorporating motion into the image reconstruction process. One of the main challenges in the practical implementation of tomographic reconstruction on a polyhedral mesh is generating the mesh geometry that is optimal for the particular imaging problem. In this work, we propose to generate the reconstruction mesh using both the projection data and the results of the finite element (FE) mechanical modeling study described below.

The most straightforward and common technique for cardiac motion compensation is motion gating. Usually, electrocardiography (ECG) is used to divide the acquired tomographic data into several temporal bins corresponding to different phase values of the quasi-periodic ECG signal. (Respiratory motion compensation is conceptually similar, although amplitude-based rather than phase-based binning and different motion tracing methods are used.) After the acquisition, each gate is reconstructed separately, providing a movie-like sequence of stationary volumes. Unfortunately, straight-forward gating, while reducing in motion artifacts, also reduces the SNR of each reconstructed image. The SNR degradation can be mitigated by temporal processing of the images during reconstruction, in other words, by switching from independent reconstruction of each time frame to a model that incorporates the frame-to-frame motion [7]. A number of different models have been proposed in the field, including describing motion using a deformable tetrahedral mesh [8]. In order to implement any of such reconstruction techniques, it is crucial to establish and validate the parameters of the modeled motion.

When identified and modeled correctly, cardiac motion as detected during a SPECT or a PET scan can provide important diagnostic information, such as an estimate of stress or strain in the heart tissue. One of the proposed techniques aimed at extracting this type of information from tomographic datasets is finite element (FE) modeling. FE computational models of the heart have been developed to gain a greater understanding of the regional stresses and strains in both the normal and pathological heart [9,10]. These models incorporate patient-specific geometry as well as the fiber-sheet microstructure of the myocardium to predict regional distributions of myocardial deformation over the heart cycle. These models can function as valuable simulation tools and have been used to study alterations in myocardial contraction resulting from cardiac disease processes [11] as well as provide estimates of wall stress which is believed to be the driving force behind remodeling of the myocardium.

[1] R. Boutchko and G. Gullberg are the Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., MS 55R0121, Berkeley, CA 94730, Email: rbuchko@lbl.gov. A. Veress is at the University of Washington, Seattle, WA

This manuscript is organized in the following manner. Section II explains the concept of a polyhedral mesh and how it is used for tomographic image reconstruction, for motion modeling and for the mechanical modeling. Section III explains our experiment and data processing algorithm. Section IV shows the initial results. Section V discusses further direction of our research, and Section VI formulates the preliminary conclusions.

II. Finite Element Meshes and Imaging

A *finite element mesh* is a combination of a number of point *node*s and linear *chords* connecting these nodes. Thus, in order to define a mesh in 3D, it is sufficient to provide an indexed set of node coordinates $\boldsymbol{r}_n$ and a set of pairs of node indices (n_1, n_2) specifying the chords. A *polyhedral mes*h is a more structured type of FE mesh, in which nodes and chords form polyhedra of various types. In addition to nodes and chords, a polyhedral mesh has two other types of elements: individual polyhedra and faces. Depending on the specific FE modeling problem, all or some of the mesh elements may be assigned different parameters such as intensities, mechanical properties, heat conducting properties, etc. In this manuscript, we use a combined geometry with hexahedral elements within the heart muscle and tetrahedral elements elsewhere in the volume of reconstruction.

A. Tetrahedral Meshes and Tomographic Reconstruction

The simplest polyhedral mesh in 3D is a mesh formed by non-overlapping tetrahedra or t*etrahedral mesh*. A tetrahedron (Figure 1(a)) is a *simplex*, the simplest non-degenerate polyhedron that can be constructed in three dimensions. In our previous work [6], we defined the mesh as:

$$\left.\begin{array}{c} K \text{ nodes } \mathbf{R}_k \text{ with assigned intensities } I_k, \\ M \text{ tetrahedra } T_m \text{ with vertices } \{m_i, m_2, m_3, m_4\}, \\ \text{where } 1 \le m_j \le K. \end{array}\right\} \quad (1)$$

Within each tetrahedron, the intensity varied linearly between the vertices, thus the image intensity distribution function was represented by a region-wise-continuous linear function. We used an analytic expression for the parallel projection of a linear distribution within a tetrahedron onto a plane to generate the system matrix and an ML-EM algorithm to reconstruct the vertex intensities I_k from projections. The optimal mesh geometry was achieved by iterative adaptive coarsening of the initially dense mesh with manually specified optimization parameters. More conservative optimization parameters for the portion of the image near the heart (the main region of interest) were selected.

A slightly different form of the image representation is used in this work for reasons to be explained in subsection 3.C. Instead of region-wise continuous approximation, we choose regionwise-constant approximation, with a constant intensity I_m assigned to each tetrahedron. This approach increases the number of unknown intensities to be reconstructed (since the number of nodes in a typical mesh is 2-3 times smaller than the number of tetrahedra), but simplifies the system matrix generation and allows defining sharper boundaries.

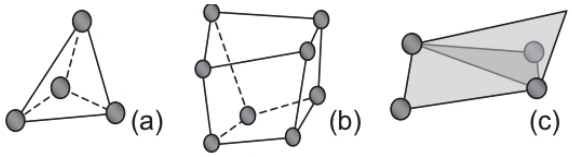


Fig. 1. (a) A tetrahedron: the 3D simplex. A non-degenerate tetrahedron is always convex, independently of the ordering of vertices, however, preservation of the vertex order is an important parameter for ensuring the mesh integrity. (b) A hexahedron. Ordering of the vertices is essential in order to define a hexahedron. A concave hexahedron is possible, however, only convex hexahedra are considered in this work. (c) Deforming a face of a hexahedron can make this face non-planar: the example shows how the vertex in the top right corner is below the plane formed by the other three vertices.

B. Hexahedral Meshes: FE Modeling of the Heart Mechanics

A hexahedron is a 3D figure with six quadrangular faces, eight vertices, and twelve chords (Fig. 1(b)). Unlike a tetrahedron, it requires that both the coordinates and the correct order of the vertices be specified. A significant complication presented by a hexahedral model is the fact that a deformation of a hexahedron can lead to one or more faces losing their planar configuration as shown in Fig. 1(c). The hexahedral mesh of the left ventricle is obtained by semi-automatic boundary segmentation of the epi- and endocardium in diastolic and systolic images and by the application of True-Grid software (www.truegrid.com). Within the mesh, the myocardium is represented as a transversely isotropic hyperelastic material [12] with fiber definitions published in the literature [13]. Subsequent mechanical analysis is performed using NIKE3D [14], a large deformation, non-linear finite element modeling software package.

C. Combined-Type Polyhedral Mesh

Our primary goal is creating a mesh that combines the hexahedral mesh H used in the mechanical model and is only defined for the myocardial tissue with the tetrahedral mesh T that occupies the rest of the region of reconstruction. Each hexahedron in H is divided into five tetrahedra as shown in Fig. 2(a). This subdivision uniquely determines the geometry of each face and, therefore, uniquely determines the subdivision of all of the interconnected elements of mesh H. Because of the complex topology of H, a consistent subdivision is not guaranteed, and we may encounter cases shown in Fig. 2(b). For the case of linear intensity approximation, this inconsistency would be irresolvable or would require further subdivision of some hexahedra causing unnecessary complexity. In the case of constant intensity approximation, we only encounter such problems if the compromised face is deformed: in that case, a small overlap between several neighboring tetrahedra may occur. Since the number of such image representation defects and the volume of each overlap is small, we ignore them in our tomographic problem.

An important component of the image representation is limiting the number of intensities to reconstruct from projections. We reduce the number of intensities to be reconstructed by bunching several mesh elements together:

- Intensities of all tetrahedra inside a single hexahedron are assumed to be identical, i.e., although tetrahedral geometry is used for constructing the system matrix, the

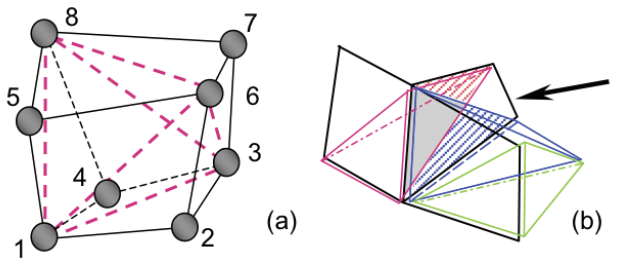


Fig. 2. (a) Splitting a hexahedron into five tetrahedra. (b) Subdividing neighboring hexahedra (black lines) into tetrahedra (colored lines, in reality should coincide with the corresponding black lines. The arrow points at the internal face divided differently from he two sides: the red and the blue tetrahedra share only the gray portion of the face, while the red-shaded and the blue-shaded corners of the respective tetrahedra do not interface. If the hexahedral face were non-planar, then the blue and the red tetrahedrons would intersect.

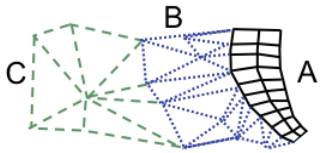


Fig. 3. Mesh deformation, triangles and quadrangles used to illustrate tetrahedra and hexahedra. Region A is a portion of *H*, with the motion model defined during mechanical modeling. Region C is the stationary portion of the tetrahedral mesh. Region B deforms driven by its boundaries (stationary boundary BC and moving boundary AB).

system matrix entries correspond to hexahedra.

- If the relative difference between the intensities of two elements *i* and *j* sharing the same face is below a pre-set threshold

$$|I_i - I_j| / \min(I_i , I_j) < \varepsilon \qquad (2)$$

for more than half of the motion frames, then the intensities of these two elements are reconstructed as one, effectively performing local image segmentation. More than two elements may be bunched in this manner, but only elements of the same type can form one segment.

Mesh deformation in *H* is computed during the mechanical modeling stage. We extend the deformation to *T* by subdividing it into three sub-regions as in Figure 3. Region A is the same as mesh *H*, where the motion is defined. Region B is the portion of the tetrahedral mesh within 5-7 cm of region B. Region C is the rest of the image. No node motion in region C is permitted. Nodes can move in region B, driven by the motion of its boundaries, the displacement of the nodes $\mathbf{R}_k$, $k \in$ B, is determined by solving the optimization problem:

$$\left.\begin{array}{c}\Sigma_{k\in B}\,|\Delta\mathbf{R}_k,| \text{ is minimized;}\\ \text{motion of AB is accommodated, BC is stationary;}\\ \text{mesh integrity is preserved (no overlaps).}\end{array}\right\} \qquad (3)$$

III. Imaging Experiment and Data Preprocessing

A. Data Acquisition

A dual head GE Millennium VG Hawkeye SPECT/CT system with high resolution parallel-hole collimators was used to image the distribution of ^{99m}Tc-sestamibi in the myocardium of patients. Acquisition was performed 1.5 hours after the injection of 20 mCi of ^{99m}Tc-sestamibi. Two detectors in L-mode rotated 90° obtaining a total of 30 projections per detector. Each 128×88 projection was binned into 8 bins of the same length, depending on the ECG phase, thus, eight projection sinograms of length N=128×88×60 were created. An X-ray CT transmission study was performed immediately prior to the SPECT scan. An average of 6×10^5 counts was acquired by each detector head for each angular view per one cardiac gate.

B. Initial Image Reconstruction

The acquired projection data were reconstructed using 20 iterations of the ML-EM algorithm with attenuation and point response corrections. The system matrix was constructed using a standard ray-tracing algorithm. The diverging ray geometry was realized by tracing multiple rays with the intensity dependence on the divergence direction established empirically (as a point-spread function) for the given scanner/collimator combination. Scanner-acquired CT images were used to compute the attenuation correction for each ray.

The initial images were reconstructed on a 128×128×88 grid with 0.442 cm voxel size. Each of the eight time frames was reconstructed separately.

C. Combined Mesh Generation

The initial hexahedral FE model was developed to study the effects of the infarct on contractile function for a 60 year old male patient with a lateral wall infarct. The node coordinates, element definitions and node order for this mesh were defined based on the SPECT images to be consistent with requirements of NIKE3D, which was used for subsequent mechanical modeling. Normal pressure loads for the cardiac cycle were used for the model analysis with [15,16]. As the result of the modeling, complete information about the node coordinates, mesh topology, and node displacements for different motion frames was generated and validated.

A tetrahedral mesh was generated in the region outside of the myocardium and connected to the hexahedral mesh described above. The nodes of the tetrahedral mesh were generated at a random location within each voxel, whereas the probability *p* of creating a node inside voxel *n* changed from 0.1 outside of the torso to 0.9 in the regions near the heart. The TetGen software package was used to generate the nodes in constrained geometry with defined faces. After the bunching process described in equation (2), a total of *M* intensities was reconstructed.

D. Motion-Compensated Image Reconstruction

Using the deformation model, a system matrix that related the mesh element intensities to the projection bins was generated. The system matrix entries inside the deformable elements were modified by the relative changes in the element volume in order to take into account the conservation of mass. (This operation was not applied to the tetrahedra inside the LV cavity since variable volume is expected there.) Ultimately, the $M\times(8N)$ system matrix was created, relating all of the projection data to *M* unknown intensities. ML-EM algorithm was used to reconstruct the intensity values.

IV. Preliminary Results

Figure 3 shows some of the preliminary results that are available at this time. In Fig. 3(a), we show the voxel-based

reconstruction of the diastolic motion frame and the schematic illustration of the hexahedral mesh generation. The actual FE mesh generated for this experiment is shown in Fig. 3(b). The results of the mechanical modeling study are analogous to the previous study described in [16].

V. Discussion and Future Work

Our future work on this project includes generating images reconstructed with and without the deformation model, and demonstrating incremental improvements to the mesh geometry and reconstructed intensities that can be achieved by iterative application of motion-compensation reconstruction and re-meshing, and local segmentation of the image.

Generation of the reliable non-rigid motion model is one of the most difficult steps of any motion-compensation methodology, therefore, the first and the most obvious result of the presented method is in recycling the motion information obtained and validated during the mechanical modeling stage. Another important result of our work is further development of the methodology of FE generation and application in tomography, specifically, combining meshes of different topologies in one problem, using mesh elements with constant intensity, and reduction of the number of reconstructed intensities by applying local segmentation (mesh element bunching). We feel that FE meshes are underutilized as data representation and post-processing tools in tomographic imaging and that further research in this field is needed. At this stage, the crucial step of the presented work (generating the LV mechanical model) is performed semi-manually. In the future, we aim to research the application of tetrahedral mesh for modeling tissue properties. A tetrahedral mesh can be generated and segmented using our currently available tools, which means that the mechanical modeling and analysis can become fully automated.

VI. Conclusion

We developed a method of incorporating the motion information and finite element mesh geometry developed during mechanical modeling of the heart motion based on the gated SPECT images to allow a cardiac motion-compensated image reconstruction method. The images are represented as intensities inside the elements of a combined hexahedral-tetrahedral deformable mesh. The mesh deformation model was derived from the gated images and validated by modeling the local stress and strain in the myocardial tissue. Local segmentation was used to decrease the number of unknown intensities and thus reduce the image noise. The modeling and motion-compensation reconstruction may be repeated iteratively improving the intensity images and optimizing the FE mesh. Further development of this method may lead to automatic mesh generation directly from projections.

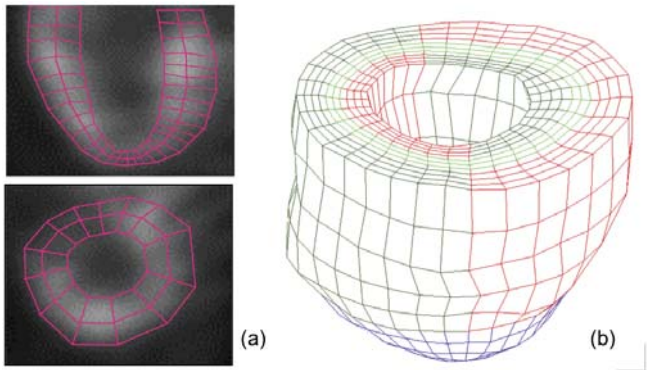


Fig. 3. (a) Processing SPECT_reconstructed images to create the hexahedral mesh. (b) The mesh used in mechanical modeling of LV motion using NIKE3D software.

Acknowledgment

The work presented in this article has been funded in part by National Institutes of Health grants R01-HL50663, R01-EB07219, and R01-EB00121 and by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] Brankov JG, Yang Y, and Wernick MN 2004 “Tomographic image reconstruction based on content-adaptive mesh model", IEEE T Med Imaging 23(2) 202-212.

[2] Brankov JG, Yang Y, and Wernick MN 2005 "Spatio-temporal processing of gated SPECT images using deformable mesh modeling" Med Phys 32(9) 2839-2849.

[3] Gonzalo R and Brankov JG 2008 "Mesh model 2D reconstruction operator for SPECT", Proc. SPIE, 6913 69132L.

[4] Sitek A, Huesman RH, and Gullberg GT 2006 "Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud" IEEE T Med Imaging 25 1172- 1179.

[5] Pereira NF and Sitek A 2010 "Evaluation Of A 3D Point Cloud Tetrahedral Reconstruction Method", Phys. Med. Biol. 55 5341-5361.

[6] Boutchko R, Sitek A. and Gullberg G., "Practical implementation of tetrahedral mesh reconstruction in emission tomography", Physics in Medicine and Biology (2nd revision).

[7] Niu X, Yang Y, Jin M, Wernick MN, King MA, "Effects of motion, attenuation, and scatter corrections on gated cardiac SPECT reconstruction", Med Phys. 2011 Dec;38(12):6571-84.

[8] Marin, Thibault, and Jovan G. Brankov. "Deformable left-ventricle mesh model for motion-compensated filtering in cardiac gated SPECT." Medical physics 37.10 (2010): 5471.

[9] McCulloch, A.D., *Cardiac Biomechanics*, in *Biomedical Engineering Handbook*. 1995, IEEE Press: Boca Raton. p. 418-439.

[10] Hunter, P., et al., *A mathematical model of cardiac anatomy*, in *Computational Biology of the Heart*. 1997, Wiley. p. 171-215.

[11] McCulloch, A.D. and R. Mazhari, *Regional myocardial mechanics: integrative computational models of flow- function relations.* Journal of Nuclear Cardiology, 2001. **8**(4): p. 506-19.

[12] Weiss, J.A., B.N. Maker, and S. Govindjee, *Finite element implementation of incompressible, transversely isotropic hyperelasticity.* Computer Methods in Applied Mechanics and Engineering, 1996. **135**: p. 107-128.

[13] Vaplon, S., J. Omens, and A. McCulloch. *Ventricular tissue adaptation associated with collagen deficiency in the osteogenesis imperfecta murine.* in *Presented at 1999 Bioengineering Conference, Big Sky, Montana*. June 16-20, 1999.

[14] Maker, B.N., R.M. Ferencz, and J.O. Hallquist, *NIKE3D: A nonlinear, implicit, three-dimensional finite element code for solid and structural mechanics.* Lawrence Livermore National Laboratory Technical Report, 1990. UCRL-MA, #105268.

[15] Veress, A.I., et al., *Incorporation of a left ventricle finite element model defining infarction into the XCAT imaging phantom.* IEEE Trans Med Imaging, 2011. **30**(4): p. 915-27.

[16] Veress, A.I., et al., Normal and pathological NCAT image and phantom data based on physiologically realistic left ventricle finite-element models. IEEE Trans Med Imaging, 2006. **25**(12): p. 1604-16.

Few-view CT Reconstruction Aided by Low-resolution Projections

Chuang Miao, Ge Wang and Hengyong Yu

***Abstract*—Traditionally, there are two strategies to reduce x-ray radiation dose: one is to reduce the x-ray flux towards each detector element, and the other is to decrease the number of measurements across a whole object to be reconstructed. Inspired by compressive sensing theory, the total variation (TV) minimization based algorithms were developed to reconstruct high-quality CT images from few-view data. However, some detailed structures may be removed by the TV-based regularization. In this paper, we propose to improve the image quality from few high resolution views aided by ultra-low-dose (low-resolution) projections. That is, projections are a combination of few-view high resolution and regular low-resolution projections. To analyze the performance of the proposed reconstruction scheme, a clinical experiment is performed. The results show the merits of the proposed methodology.**

***Index Terms*—Computed tomography (CT), low-dose CT, compressive sensing (CS), few-view data, low-dose/low-resolution data.**

I. Introduction

X-RAY computed tomography (CT) has been extensively used in clinics as a primary diagnostic imaging modality since its invention. However, the x-ray radiation is harmful which may induce genetic, cancerous, and other diseases [1-3]. Because of the concerns the x-ray radiation to the patients, minimizing the radiation dose has been one of the major endeavors in the current computed tomography (CT) examinations [3, 4]. As a result, the well-known ALARA (as low as reasonably achievable) principle is well accepted to avoid excessive radiation dose in the medical community. Therefore, it is highly desirable to reduce the radiation dose while maintaining the clinically acceptable image quality.

There are two strategies to reduce radiation dose: one is to reduce the x-ray flux towards each detector element, and the other is to decrease the number of x-ray projections across a whole object to be reconstructed. The former is usually implemented by reducing the x-ray dose to lower mAs levels or tube current modulation. Nonetheless, this approach will result in an insufficient number of x-ray photons detected at the detector and hence elevate the quantum noise level on the projections. As a consequence, the quality of the CT images reconstructed from a conventional filtered backprojection (FBP) algorithm [5] will be degraded by the noisy projection data. The latter necessarily produces insufficient data measurement suffering from serious streaking artifacts in the reconstructed CT images, since the FBP algorithm requires that the sampling rate of projections should satisfy the Shannon sampling theorem [6].

Recently, the compressive sensing (CS) theory shows that a compressible signal can be accurately reconstructed from samples far less than what is required by the Shannon/Nyquist sampling theorem [7, 8]. The key is to enforce some additional constraints to select the "best" candidate. It has been shown that finding the candidate with the minimum l_1 norm, which is closely related to the total variation (TV) minimization in a number of imaging cases assuming a piecewise constant image model [9], is the most reasonable choice [7, 8]. In such an approach, the regularization term of a TV form is minimized under the data fidelity constraints posed by the x-ray projections. This approach has also been widely applied in many other imaging modalities [10-13] and its efficacy has been improved by incorporating prior information [14, 15].

In this paper, we improve the TV-minimization based image quality for few high-resolution (HR) views CT reconstruction by incorporating ultra-low-dose low-resolution (LR) projections. Extensive clinical experiments will be performed to quantitatively and qualitatively evaluate the proposed data acquisition protocol and reconstruction algorithm. The rest of this paper is organized as follows. In Section II, we will describe the reconstruction approach and its implementation. In Section III, we will report representative results from clinically acquired datasets, and quantify the performance of our proposed methodology. Finally, in Section IV we will conclude the paper.

II. Algorithm Development

A. CT Imaging System Model

Many imaging systems, such as CT scanners, can be modeled by the following linear equations [16]:

$$Wf = p, \tag{1}$$

where $p \in P$ represents projection data, $f \in F$ represents an unknown image, and the non-zero matrix $W : F \to P$ is a projection operator. Although equation (1) is not new, it should be pointed out the projection data p includes measurements from both high-resolution few-view projections and low-resolution regular projections, which make it different from the existing few-view reconstruction methods in the CT field.

This work was supported in part by NIH Grant EB011785, NSF Grant CBET-1149679 and NSF Grant DMS-1210967.

C. Miao and H.Y. Yu are with the Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157 USA. (e-mail: cmiao@wakehealth.edu, hengyong-yu@ieee.org).

G. Wang is with the Biomedical Imaging Center, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, USA. (e-mail: ge-wang@ieee.org).

B. Minimization Algorithm

A CT reconstruction problem is formulated as to reconstruct the unknown vector f based on the system matrix W and the observation vector p. When it comes to an undersampled problem where the row size is smaller than the column size of W, the equation (1) becomes underdetermined and there exist infinitely solutions. However, if the gradient of the image f is sparse, it can be solved by minimizing the cost function with a TV regularization term:

$$f = \arg\min_f \|Wf - p\|_2^2 + \mu \cdot J(f) \qquad (2)$$

Where $\|\cdot\|_n$ denotes the l_n norm.

The solution of equation (2) can be determined iteratively by an alternative minimization method with two major steps. In equation (2), the first term is known as a data fidelity term. The ordered-subset simultaneous algebraic reconstruction technique (OS-SART), and some other algorithms can be used to fulfill this step to get a rough solution. For example, we can use the OS-SART method [20] to update the reconstructed intermediate image:

$$f_n^{(k+1)} = f_n^{(k)} + \lambda_k \sum_{m \epsilon \phi_l} \frac{\omega_{mn}}{\sum_{m' \epsilon \phi_l} \omega_{m'n}} \frac{p_m - \tilde{p_m}}{W_{m+}}, \quad k = 0, 1, 2, \ldots, \qquad (3)$$

where k indicates the number of iteration, λ_k is relaxation parameter (in our implementations, we set λ_k to be 1 for simplicity), ϕ_l represents the set of ray indexes in the l^{th} subset $(l = k \bmod N_\phi + 1 \epsilon \{1, 2, \ldots, N_\phi\}$, N_ϕ is the total number of subsets), $\tilde{p_m} = \sum_{n=1}^{N} \omega_{mn} f_n^{(k)}$ and $W_{m+} = \sum_{n=1}^{N} \omega_{mn}$, which can be viewed as the normalized length of the m^{th} ray path through the object. When all equations in (1) are used in one iteration step, it becomes SART [20].

The second term in equation (2) is a regularization term, which is chosen to be a TV norm. Specifically, the TV regularization term of image $f = (f_{i,j}) \in R^I \times R^J$ can be defined as [21, 22]

$$J(f) = \|f\|_{TV} = \sum_{i,j} \sqrt{(f_{i,j} - f_{i-1,j})^2 + (f_{i,j} - f_{i,j-1})^2}\,. \qquad (4)$$

A scalar μ is used to adjust the relative weights between the data fidelity term and the regularization term, which was chosen empirically to yield a good reconstruction in this work. In order to implement the TV minimization, we use the steepest descent search algorithm (TVM-SD) [21].

The weighted method in the fast iterative shrinkage-thresholding algorithm [23] was employed to accelerate the convergence of the OS-SART. The pseudo-code of the algorithm in this paper is summarized in Table I. In each of the main loop, the OS-SART is used to enforce data consistency, the TVM-SD is used to minimize the image TV and the weighting is to speed up the convergence. The three steps are applied iteratively until the stopping criteria are satisfied, which are a maximum number of iterations and an error threshold in the projection domain. The stopping criteria for iterative reconstruction were theoretically and experimentally studied before. The optimal rules for the proposed algorithm are open to be investigated, and preferably in a task-specific fashion. Given the pilot nature of this technical contribution, we will perform neither mathematical analysis nor systematic simulation but we will follow up along this direction in a future study.

TABLE I
PSEUDO-CODE OF THE ALGORITHM

Choose parameter μ.

Initialize iteration number $k = 0$ and f^k

Step 0. Take $h^1 = y^1 = f^0 \in R^n$, $t_1 = 1$.

Step k ($k \geq 1$). Compute

OS-SART: $f^k = \text{OS-SART}(y^k)$

TVM-SD: $h^k = \text{TVM-SD}(f^k)$

Fast Weighted: $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$

$$y^{k+1} = h^k + \left(\frac{t_k - 1}{t_{k+1}}\right)(h^k - h^{k-1})$$

Repeat step k until the stopping criteria is satisfied.

III. Experiment Results

A. Clinical Study

1) Data Acquisition:

The proposed few-view reconstruction method was evaluated by a group of existing raw projections collected in a cardiac perfusion CT study for other purposes. The patient was scanned by a state-of-the-art GE discovery CT750 HD scanner to examine the coronary artery. After appropriate pre-processing, a fan-beam sinogram was used in the study. The radius of the scanning trajectory was 53.852 cm. Over a 360 scanning range, 2200 projections were uniformly acquired. For each projection, 888 detector elements were equi-angularly distributed over an FOV of 24.92 cm in radius. We combined two, four and eight detector elements into one to simulate three sets of low-resolution projections with 444, 222 and 111 detector elements, respectively.

Then we reconstructed images from few-view high-resolution projections aided by low-resolution datasets. For different projection datasets, the same parameters, such as the relaxation parameter and the number of views in each subset, were used in the OS-SART algorithm. The initial images were set as zero and the size of all the subsets of OS-SART was set to 11. The reconstructed images included 289×353 pixels. Each pixel covered an area of $457.7 \times 457.7\ \mu m^2$, which is comparable to resolution of the original projection data with 888 detector elements.

2) High Resolution Few-view Test

Reducing the number of projections is important to reduce radiation dose. The HR projections were down-sampled from 2200 views to 583, 396 and 209, respectively. Each down-sampled subset of projections were interpolated uniformly around 360 degrees. Images were reconstructed from the down-sampled dataset using the OS-SART with TV regularization. The reconstruction from 2200 HR projections using OS-SART was selected as the reference image for analysis.

Qualitative Indexes Evaluation: the image quality of the reconstructed images was visually evaluated. As shown in Fig.

1, the first image in was reconstructed from all the HR projections (2200 views).. The second to fourth images were reconstructed from down-sampled high-resolution projections (583, 396 and 209 views, respectively). It can be seen that the very good image was reconstructed from 583 views. The image reconstructed from 396 views was blurred in some finer details such as the region indicated by an arrow 'A'. The image reconstructed from 209 views was more blurred and even missing some finer details such as the details around the region indicated by the arrow 'A'. This image was also blocky in the region indicated by the arrow 'B'.

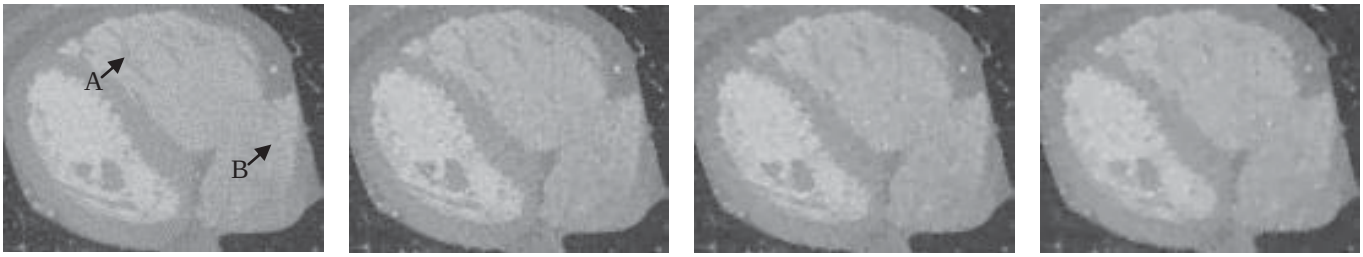


Fig. 1. Magnified cardiac regions of the reconstructed images from high-resolution projections. The left image was reconstructed OS-SART method from all the projections(2200 views), and the 2nd to 4th images were reconstructed by the OS-SART & TV regularization method from down-sampled 583, 396 and 209 projections, respectively. The display window was [-100 1000] HU.

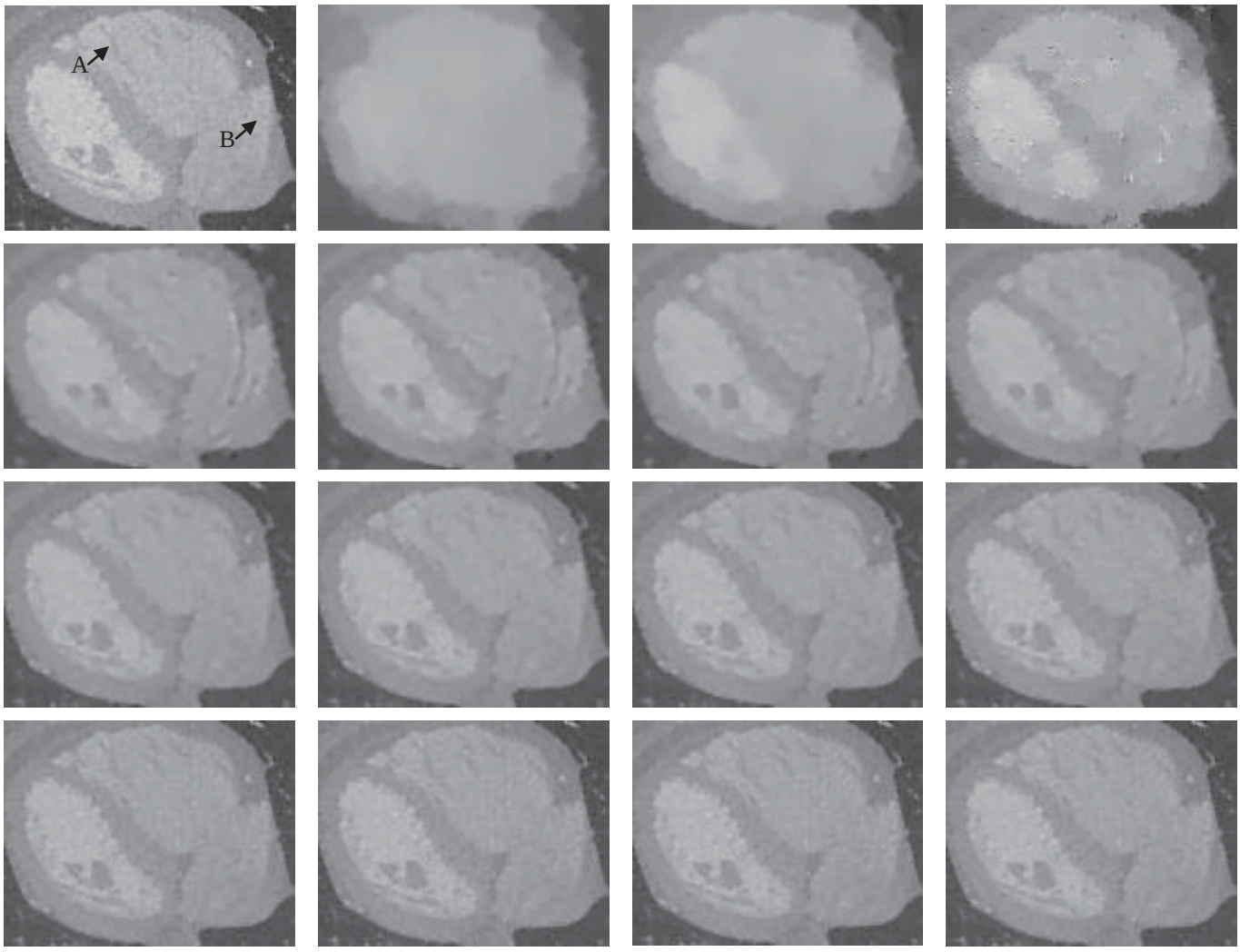


Fig. 2. Magnified cardiac regions of the reconstructed images from 583 projections. In the first column, from left to right, the images were reconstructed from projections with 888(HR), 444, 222 and 111 detector elements, respectively. The images in the second to fourth columns, from left to right the first row images were reconstructed only from 11, 22 and 33 HR projections, respectively. The images in the second to fourth columns, from the second to fourth rows, were reconstructed from LR projections corresponding images in the first column by combining 11, 22 and 33 HR projections, respectively. The display window is [-100 1000] HU.

Quantitative Indexes Evaluation: The reconstructed images were quantitatively evaluated using the image quality assessment index for structural similarity (SSIM). SSIM was designed to improve the traditional methods such as the peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have been shown to be inconsistent with human visual

perception [14]. The SSIM index is a well-established metric to measure the image quality relative to the reference image. If the measured image is exactly same with the reference image, the SSIM index will be equal to 1. With the OS-SART reconstruction from all the HR projections as the reference, we evaluated the reconstructed images in terms of SSIM as listed in Table II. We can see that the SSIM values of the reconstructed images from 396 views to 583 views don't change too much.

TABLE II
SSIM comparison of the reconstructed images from down-sampled high-resolution projections

Views	583	396	209
SSIM	0.8270	0.7964	0.6918

From the above analysis, it seems that 583 views are sufficient to obtain good enough image quality. We will select 583 views as an example to analyze the proposed approach in the mixed resolution (MR) reconstruction experiment.

3) Mixed-resolution (MR) Experiment

In order to evaluate the proposed reconstruction technique, we performed the MR experiments. We selected the same 583 projection views in each LR projection data (projection data with 444, 222 and 111 detector elements). Then we replace some of the LR projection data with the same number of HR projection data in the same acquisition angle. Here the HR projection number was set as 11, 22 and 33. Then the images were reconstructed by the OS-SART and TV regularization method. To demonstrate the merits of the proposed approach, images were also reconstructed images only from the few-view HR projections (11, 22 and 33 views). The results were shown in Fig. 2.

Quantitative Indexes Evaluation: The SSIM values were computed. The reference image was selected as the reconstructed image from 583 HR projections. As indicated in Table III, the image quality of few-view reconstruction from HR projections (second row) was significantly improved by incorporating low-resolution projections. For the reconstructed images from the projection data with same number of LR detector elements (third to fifth rows in the table), the image quality increase as the number of HR projection views increase.

TABLE III
SSIM values of the reconstructed images with the result from 583 HR projections as the reference. The numbers in first column represent the number of detector elements in the projections. The numbers in the first row represent the number of HR projections used in reconstruction. The second row values are the SSIM values of the images reconstructed only from few-view HR projections without LR projection constraints.

	0 HR	11 HR	22 HR	33 HR
No LR		0.1328	0.1820	0.2116
LR(111)	0.5510	0.5591	0.5677	0.5760
LR(222)	0.7167	0.7222	0.7290	0.7357
LR(444)	0.8800	0.8824	0.8841	0.8859

IV. Discussion and Conclusion

In conclusion, we have proposed and evaluated a few-view reconstruction strategy for high-resolution images aided by low-resolution projections. Our results confirm that the proposed method can significantly increase the image quality. In the near future, we will further investigate the proposed technique in in biomedical applications, experimentally establish the relationship between the reconstructed image quality, the high-resolution and low-resolution view numbers, as well as radiation dose, which will result in practical ultra-low-dose CT protocols.

V. References

1. Brenner, D.J., et al., *Estimated risks of radiation-induced fatal cancer from pediatric CT.* A. J. of Roentgenology, 2001. **176**(2): p. 289-296.
2. Berrington de Gonzalez, A. and S. Darby, *Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries.* Lancet, 2004. **363**(9406): p. 345-51.
3. Brenner, D.J. and E.J. Hall, *Computed tomography--an increasing source of radiation exposure.* N Engl J Med, 2007. **357**(22): p. 2277-84.
4. Einstein, A.J., M.J. Henzlova, and S. Rajagopalan, *Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography.* Jama-Journal of the American Medical Association, 2007. **298**(3): p. 317-323.
5. Deans, S.R., *The Radon Transform and Some of its Applications.* New York: Wiley), 1983.
6. Jerri, A.J., *The Shannon sampling theorem—Its various extensions and applications: A tutorial review.* Proceedings of the IEEE 1977. **65**(11): p. 1565-1596.
7. Donoho, D.L., *Compressed sensing.* IEEE Transactions on Information Theory, 2006. **52**(4): p. 1289-1306.
8. Candes, E.J.R., J. Tao, T., *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information.* IEEE Transactions on Information Theory, 2006. **52**(2): p. 489- 509.
9. Rudin, L.I., *et al*., *Nonlinear total variation based noise removal algorithms.* Physica D: Nonlinear Phenomena, 1992. **60**(1–4): p. 259-268.
10. Ma S.Q. Yin W. Zhang Y. Chakraborty, A., *An efficient algorithm for compressed MR imaging using total variation and wavelets.* IEEE Conference on Computer Vision and Pattern Recognition 2008: p. 1-8.
11. Persson, M., D. Bone, and H. Elmqvist, *Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography.* Phys Med Biol, 2001. **46**(3): p. 853-66.
12. Sidky, E.Y., et al., *Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms.* Medical Physics, 2009. **36**(11): p. 4920-4932.
13. Song, J., et al., *Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT.* Medical Physics, 2007. **34**(11): p. 4476-4483.
14. Chen, G.H., *et al., PICCS: A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets.* Medical Physics, 2008. **35**(2): p. 660-663.
15. S. Cho, E.P., E.Y. Sidky et al., *Prior-image-based few-view cone beam CT for applications to daily scan in image-guided radiation therapy: preliminary study.* Proc SPIE – Int Soc Opt Eng, 2009.
16. Kak, A.C. and M. Slaney, *Principles of Computerized Tomographic Imaging (New York: IEEE Press).* 1999.
17. De Man, B. and S. Basu, *Distance-driven projection and backprojection.* IEEE Nuclear Science Symp. Medical Imaging Conf. (Norfolk), 2002.
18. De Man, B. and S. Basu, *3D distance-driven projection and backprojection.* Proc. 7th Int. Conf. on Fully 3D Reconstruction in Radiology and Nuclear Medicine (Saint Malo), 2003.
19. Yu, H. and G. Wang, *Finite detector based projection model for high spatial resolution.* J. of x-ray Sci. and Tech., 2012. **20**(2): p. 229-38.
20. Wang, G. and M. Jiang, *Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART).* J. of x-ray Sci. and Tech., 2004. **12**(3): p. 169-177.
21. Yu, H. and G. Wang, *Compressed sensing based interior tomography.* Physics in Medicine and Biology, 2009. **54**(9): p. 2791-2805.
22. Sidky, E.Y., C.M. Kao, and X.H. Pan, *Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT.* Journal of X-Ray Science and Technology, 2006. **14**(2): p. 119-139.
23. Beck, A. and M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.* SIAM Journal on Imaging Sciences, 2009. **2**(1): p. 183-202.

Cost-Effectiveness of Noise Reduction in Iterative Reconstruction Using Ordered Subset SIRT with Total Variation Regularization

Shaojie Tang and Xiangyang Tang

***Abstract*—It is well-known that the radiation dose and spatial resolution in CT with filtered backprojection algorithm for image reconstruction observes the 4th-power law. To break down this dose-demanding law, iterative image reconstruction with various regularization schemes has recently become a subject of active research. We propose and implement an ordered-subset Simultaneous Iterative Reconstruction Technique (OS-SIRT) with Total Variation (TV) regularization for advanced clinical applications that require low radiation dose and high spatial resolution simultaneously. In this work, using a performance phantom, we conduct a quantitative study to investigate the cost-effectiveness (reduction of noise vs. degradation of spatial resolution) in iterative image reconstruction using the OS-SIRT combined with TV regularization. In addition, we explore the strategy of recovering the degradation in spatial resolution caused by the TV regularization in the OS-SIRT algorithm. It is believed that the preliminary results presented in this paper can provide information adding to an insightful understanding of the cost-effectiveness of OS-SIRT algorithm with TV regularization in noise reduction and spatial resolution maintenance.**

***Index Terms*—Filtered backprojection, Ordered-subset, Simultaneous iterative reconstruction technique, Total variation, Spatial resolution, Noise**

I. Introduction

In vast majority of clinical CT scanners, images are reconstructed by the filtered backprojection (FBP) algorithm [1], in which the required radiation dose is proportional to the 4th-power of spatial resolution [2]. With an increasing awareness of the radiation dose associated with x-ray CT and its potential biological consequences [3], a break-down of this 4th-power law through algorithmic solutions has become a subject of active research [4]. In the image reconstruction using FBP algorithms, linear low-pass filtering or smoothing is an effective way to reduce radiation dose, but always at the cost of spatial resolution degradation in observation of the 4th-power law [2]. Another approach is the image reconstruction using iterative algorithms combined with various regularization schemes [4], in which the noise is substantially reduced while the spatial resolution is maintained significantly better than that in its FBP counterpart [1]. Numerous iterative image reconstruction algorithms differing in the regularization schemes have been proposed and most of them have shown great results. However, we believe that the cost-effectiveness of iterative image reconstruction combined with various regularization schemes in reducing noise and maintaining spatial resolution deserves a further and in-depth scrutiny.

Using a performance phantom and the ordered-subset Simultaneous Iterative Reconstruction Technique (OS-SIRT) [5,6] combined with Total Variation (TV) regularization scheme [7], we carry out a quantitative investigation in this work. The primary purpose is to answer the following questions: (*i*) how the trade-off between noise and spatial resolution in the reconstruction using the OS-SIRT algorithm without regularization scheme is different from that using the FBP algorithm; (*ii*) how the regularization scheme, e.g., total variation, can be a game changer in the trade-off between noise and spatial resolution in image reconstruction using the SIRT algorithms; and (*iii*) if a boosting of high frequency components in the projection domain prior to the iterative reconstruction can recover the degradation in spatial resolution caused by the regularization.

Shaojie Tang is with the Department of Radiology & Imaging Sciences, Emory University School of Medicine, Atlanta, GA, 30322, USA, and the School of Automation, Xi'an University of Posts & Telecommunications, Xi'an, Shaanxi, 710121, China, (e-mails: shaojie.tang@emory.edu; tangshaojie@xupt.edu.cn)

Xiangyang Tang is with the Department of Radiology & Imaging Science, Emory University School of Medicine, Atlanta, GA, 30322, USA, (corresponding author, e-mail: xiangyang.tang@emory.edu).

II. Theorem Background

A. Geometry of Data Acquisition and Image Reconstruction

The equi-angular fan-beam geometry is shown in Fig.1, where *O-xy* represents the Cartesian coordinate system, and S is the focal spot of x-ray source. $P(x, y)$ denotes a point within the object to be imaged. The ray emanating from focal spot S and passing through $P(x, y)$ is determined by its view angle η and fan angle γ. Then, the circular source trajectory is expressed as,

$$ST(\eta) = (R\sin\eta, R\cos\eta), \eta \in [0, 2\pi). \quad (1)$$

B. Image Reconstruction by FBP

The FBP reconstruction algorithm in the equi-angular fan-beam geometry can be expressed as [1],

$$f(x, y) = \frac{1}{2}\int_0^{2\pi} g_h(\eta, \gamma)\, d\eta, \quad (2)$$

where

$$g_h(\eta,\gamma) = g(\eta,\gamma) \otimes h(\gamma), \tag{3}$$

$$g(\eta,\gamma) = \boldsymbol{D}f(\eta,\gamma), \tag{4}$$

$h(\gamma)$ is the ramp filter kernel expressed in the equi-angular fan-beam geometry, and $\boldsymbol{D}$ represents the projection operator. In practice, the spatial resolution of the image reconstructed by the FBP algorithm can be tuned by the rolling-off and cutting-off parameters of a window function, which is equivalent to a linear low-pass filtering or smoothing, applied on the ramp filter kernel $h(\gamma)$ in the frequency domain.

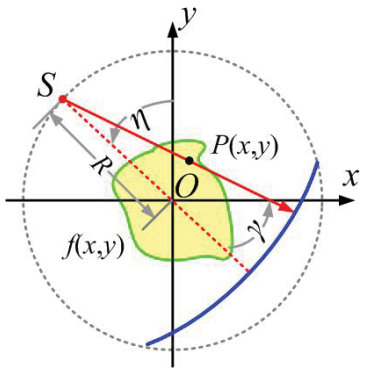


Fig.1. A diagram showing the equi-angular fan-beam geometry of data acquisition and image reconstruction.

C. Image Reconstruction by SIRT

The SIRT algorithm in the equi-angular fan-beam geometry can be expressed as [5],

$$f^{(k+1)}(x,y) = f^{(k)}(x,y) + \omega \frac{\sum_{(\eta,\gamma)} \frac{a(x,y;\eta,\gamma)\left(g(\eta,\gamma) - \sum_{(x,y)} a(x,y;\eta,\gamma) f^{(k)}(x,y)\right)}{\sum_{(x,y)} a(x,y;\eta,\gamma)}}{\sum_{(\eta,\gamma)} a(x,y;\eta,\gamma)}, \tag{5}$$

where ω denotes the relaxation parameter ranging in (0, 2), and $a(x, y; \eta, \gamma)$ is the normalized intersection area of the strip determined by the detector cell at (η, γ) and the image pixel at (x, y) [8]. The SIRT algorithm given in eq. (5) is essentially the gradient descent scheme to minimize the objective function [9],

$$\Phi_{SIRT}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{W}\boldsymbol{f}\|_W^2 = \langle W(\boldsymbol{g} - \boldsymbol{W}\boldsymbol{f}), \boldsymbol{g} - \boldsymbol{W}\boldsymbol{f} \rangle, \tag{6}$$

where the weighting matrix W is diagonal with its element $W[i, i] = W[(\eta, \gamma), (\eta, \gamma)] = 1/\Sigma_{(x,y)} a(x, y; \eta, \gamma)$, and $\langle \bullet, \bullet \rangle$ denotes an inner product in the projection domain.

D. Image Reconstruction by SIRT with TV regularization

In order to investigate the tradeoff between noise and spatial resolution in the SIRT iterative algorithm [5], the TV regularization is employed [7], in which an unconstrained minimization strategy is adopted [10]. The objective function is designed by linearly combining the weighted L-2 norm corresponding to the SIRT reconstruction [9] and the L-1 norm of image gradient corresponding to the TV regularization [7],

$$\begin{aligned}\Phi_{SIRT-TV}(\boldsymbol{f}) &= \Phi_{SIRT}(\boldsymbol{f}) + \lambda \Phi_{TV}(\boldsymbol{f}) \\ &= \|\boldsymbol{g} - \boldsymbol{W}\boldsymbol{f}\|_W^2 + \lambda \|\boldsymbol{f}\|_{TV},\end{aligned} \tag{7}$$

$$\|\boldsymbol{f}\|_{TV} = \|\nabla \boldsymbol{f}\|_1, \tag{8}$$

where $\|\bullet\|_1$ and ∇ denote the L-1 norm and gradient operations, respectively, and parameter λ is used to adjust the contribution to the reconstruction from the TV regularization.

Unfortunately, the objective function defined above is not quadratic even though it is convex. Similar to the strategy exercised in [11], the TV part of the objective function is substituted by a sequence of quadratic ones,

$$\|\nabla \boldsymbol{f}\|_1 \le \|\nabla \boldsymbol{f}\|_2^2 \Big/ \left(2\left|\nabla \boldsymbol{f}^{(k)}\right|\right). \tag{9}$$

Consequently, the original objective function given in eqs. (7) and (8) becomes,

$$\tilde{\Phi}_{SIRT-TV}^{(k)}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{W}\boldsymbol{f}\|_W^2 + \lambda \|\nabla \boldsymbol{f}\|_2^2 \Big/ \left(2\left|\nabla \boldsymbol{f}^{(k)}\right|\right). \tag{10}$$

In such a way, the following relationships hold,

$$\tilde{\Phi}_{SIRT-TV}^{(k)}(\boldsymbol{f}) \ge \Phi_{SIRT-TV}(\boldsymbol{f}), \tag{11}$$

$$\tilde{\Phi}_{SIRT-TV}^{(k)}\left(\boldsymbol{f}^{(k)}\right) = \Phi_{SIRT-TV}\left(\boldsymbol{f}^{(k)}\right). \tag{12}$$

With the mathematical treatment detailed above, the SIRT algorithm with TV regularization can be expressed as

$$f^{(k+1)}(x,y) = f^{(k)}(x,y) + \omega \frac{\sum_{(\eta,\gamma)} \frac{a(x,y;\eta,\gamma)\left(g(\eta,\gamma) - \sum_{(x,y)} a(x,y;\eta,\gamma) f^{(k)}(x,y)\right)}{\sum_{(x,y)} a(x,y;\eta,\gamma)}}{\sum_{(\eta,\gamma)} a(x,y;\eta,\gamma)}, + \omega\lambda \Delta f(x,y) \Big/ \left|\nabla f^{(k)}(x,y)\right|, \tag{13}$$

In order to speed up the convergence of the SIRT reconstruction with TV regularization, ordered-set technique is employed in a way similar to that in [6]. In this work, the number of ordered sets is empirically determined as 40.

E. SIRT with Boosting of High Frequency in Projection Data

To recover the spatial resolution degraded by the TV regularization in the SIRT reconstruction, a boosting of high frequency components is carried out in the projection domain

$$\tilde{g}(\eta,\gamma) = -\alpha g(\eta,\gamma-1) + (1+2\alpha) g(\eta,\gamma) - \alpha g(\eta,\gamma+1), \tag{14}$$

where parameter α is used to adjust the intensity of boosting, and is empirically set as 0.05 in this work.

III. Numerical Evaluation

A performance phantom with its target deployment detailed in Fig. 2 is utilized to conduct the simulation study, in which the detector array for projection data acquisition is cylindrical. The distance from the x-ray focal spot to the axis of rotation (R) is 541.0 *mm*. The full fan-angle of x-ray beam is 54.89°, spanned by 888 detector cells. A total of 1,160 projection views are

uniformly acquired over angular range [0, 2π), i.e., a full scan. The matrix dimension of reconstructed transaxial images is 1,024 × 1,024, and the reconstruction field of view (FOV) is 240 *mm*, leading to image voxel size 0.2344 × 0.2344 mm^2. In the data acquisition, the x-ray techniques are 120 *kVp* and 600 *mAs*. Both quantum and electronic noise are modeled to observe the Poisson and Gaussian distributions, respectively, in a way similar to that in [12].

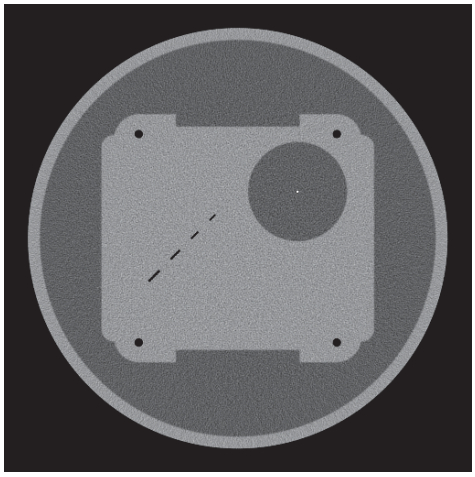


Figure 2. A transverse image of the performance phantom used to investigate the trade-off between noise and spatial resolution, in which a 50 μm tungsten wire and its surrounding area filled with water are used for MTF and noise measurement, respectively.

A. *Images Reconstructed by FBP Algorithm*

A number of rolling-off and cutting-off parameters in frequency domain are used to investigate the trade-off between noise and spatial resolution in FBP reconstruction. Due to space limitation, only the two cases corresponding to the standard case and an extremely boosted case in the spatial resolution are shown in Fig. 2. As expected, the noise increases dramatically with increasing spatial resolution, as evidenced by the sharper wire in Fig. 3 (a) and (b) and the modulated transferring function (MTF) presented in Fig. 3 (a') and (b').

B. *Images Reconstructed by OS-SIRT with TV Regularization*

The transverse images reconstructed by the OS-SIRT algorithm without (λ = 0.0) and with TV regularization (λ = 10^{-6}, 5.5×10^{-6} and 10^{-5}) are presented in Fig. 4, in which all images are acquired at the 50th iteration. A visual inspection of Fig. 4 (a – d) shows that, the TV regularization in the OS-SIRT algorithm reduces noise substantially, and degrades the spatial resolution too, but to an extent that is significantly less than the linear low-pass filtering or smoothing in an FBP reconstruction.

C. *Evaluation of Cost-Effectiveness over Algorithms*

Inclusively, the behavior of noise (gauged by the standard deviation (SD) in a water area surrounding the tungsten wire) and spatial resolution (MTF at 5%) in the image reconstructions by the OS-SIRT, OS-SIRT with TV regularization and its combination with the projection domain boosting are plotted in Fig. 5, while the quantitative measurement results are itemized in Table I. Served as a baseline, the profile in green color corresponding to the FBP reconstruction is also presented. Noted that, the metric of vertical axis of Fig. 5 is noise, and thus the profile corresponding to FBP reconstruction observes the 2nd-power law [13].

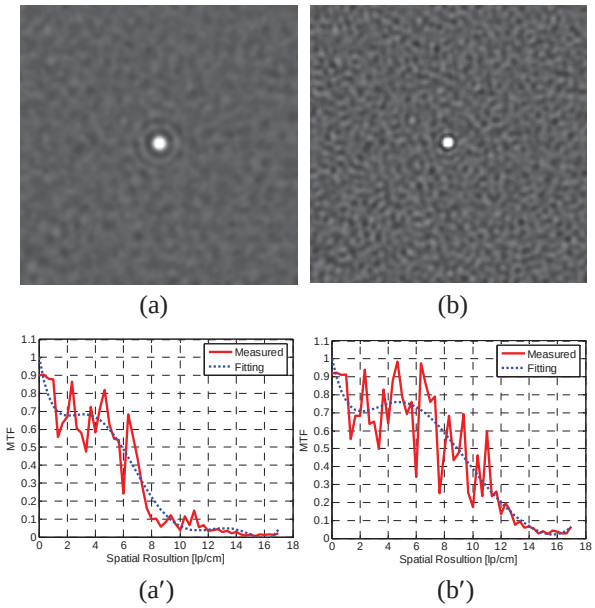


Figure 3. The transverse images of the performance phantom reconstructed by FBP algorithm with the standard (a) and extremely boosted (b) spatial resolution and their corresponding MTFs (a′ & b′).

IV. DISCUSSIONS AND CONCLUSIONS

Using a performance phantom, we carry out a quantitative study to investigate the cost-effectiveness (reduction of noise vs. degradation of spatial resolution) of the OS-SIRT with TV regularization scheme in this work. In addition, we explore the strategy of recovering the degradation in spatial resolution caused by the regularization scheme in the OS-SIRT algorithm. Referring to Fig. 5, a summary of observations is given below.

As anticipated, if each detector cell is a two-dimensional square, the noise (dose) and spatial resolution in the image reconstructed by the FBP algorithm observes the 2nd-power (4th-power) law. The accountable for this poor cost-effectiveness is the ramp filter kernel [2,13]. In the images reconstructed by the OS-SIRT without TV regularization, both noise and spatial resolution increase with the number of iteration increases. Initially, the noise increase in a manner that is less aggressive than the FBP algorithm. However, if the number of iteration exceeds a threshold, the noise increases with spatial resolution in a manner that is even more aggressive than the FBP algorithm, i.e., there exist an inflection point in the cost-effectiveness of noise reduction in the image reconstructed by the OS-SIRT algorithm without regularization. Moreover, as illustrated by the corresponding profile, the OS-SIRT with TV regularization at λ = 10^{-6} cannot suppress

the noise very well with increasing spatial resolution, but avoids entering the over-aggressive region while the number of iteration is over the threshold. However, at $\lambda = 10^{-5}$, the spatial resolution in the image reconstructed by the OS-SIRT algorithm increases with the number of iteration, but the noise does not. This demonstrates that an adequate TV regularization can keep the noise stable while the spatial resolution increase with the number of iteration and is exactly what is desired.

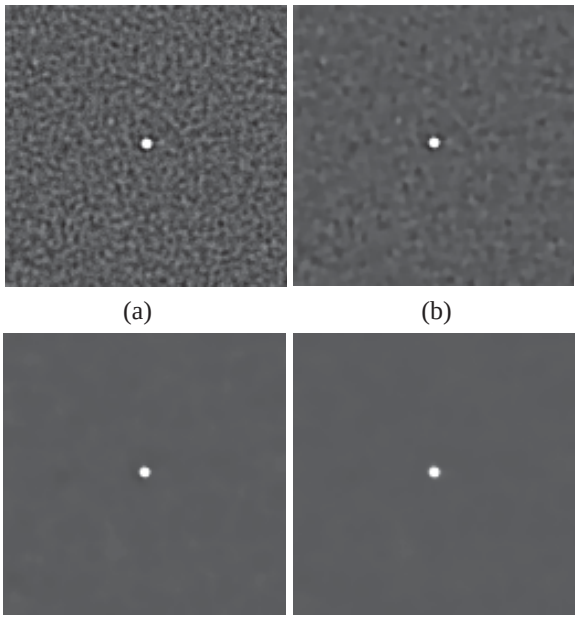


Fig.4. The zoom-in images of the performance phantom reconstructed by the SIRT algorithm with TV regularization at λ equal to 0, 10^{-6}, 5.5×10^{-6} and 10^{-5}.

TABLE I. THE SPATIAL RESOLUTION (5%MTF) AND NOISE (SD) OF THE FBP ANALYTIC AND THE SIRT-TV ITERATIVE RECONSTRUTIONS.

Reconstruction algorithm		Spatial resolution (5%MTF [lp/cm])	Noise (SD [10^{-4}/mm])
FBP	standard	10.2882	2.2699
FBP	boosted	14.4209	4.6403
SIRT-TV λ=0	no boost	16.8224	6.1274
	boosted	16.8194	7.0967
SIRT-TV $\lambda=10^{-6}$	no boost	14.8351	2.5482
	boosted	15.2228	3.2377
SIRT-TV $\lambda=5.5\times10^{-6}$	no boost	14.2521	0.4683
	boosted	14.5470	0.5142
SIRT-TV $\lambda=10^{-5}$	no boost	13.5325	0.3070
	boosted	14.0657	0.3275

It should be pointed out that, since the OS-SIRT with TV regularization is basically a nonlinear algorithm, the denotation of its spatial resolution by the 5% MTF is quite heuristic, and the ultimate assessment of image quality should be carried out via human or model observer study. In addition, it should be noted that, even though in a much tamer manner than the low-pass filtering or smoothing in the FBP reconstruction, the TV regularization in the OS-SIRT degrades spatial resolution compared to the OS-SIRT without regularization. However, as demonstrated by the thin tip at the right end of the profiles corresponding to $\lambda = 10^{-5}$, the spatial resolution degradation caused by TV regularization can be recovered by boosting the high frequency of projection data before iterative reconstruction. A quantitative study is being conducted to investigate how much of the degraded spatial resolution can be recovered by such a projection domain boosting strategy, and the results will be promptly published in the near future.

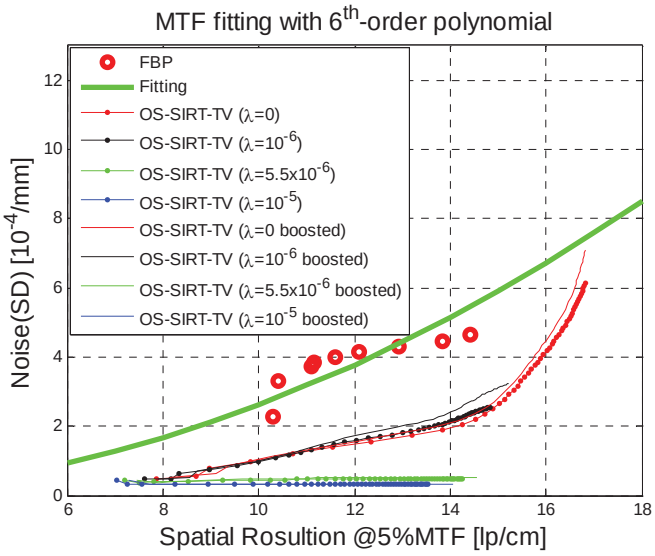


Fig.5. Variation of noise over spatial resolution denoted by 5%MTF in image reconstructions using FBP and OS-SIRT with TV regularization and projection domain boosting of high frequency (boosted), wherein the solid green curve is fitted by a 2nd-power function [11].

REFERENCES

[1] A. C. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, Chapter 3: Algorithms for reconstruction with nondiffracting sources, IEEE Press, pp.49, 1988.

[2] H. H. Barrett and W. Swindell, Radiological Imaging: The Theory of Image Formation, Detection, and Processing, Academic Press, 1981.

[3] D. J. Brenner and E. J. Hall, "Computed tomography – an increasing source of radiation exposure," N. Engl. J. Med., vol.357, pp.2277, 2007.

[4] C. Hofmann, M. Knaup and M. Kachelrieß, "Effects of Ray-Modeling: Simulation Study," The International Conference on Image Formation in X-ray Computed Tomography, pp.17, 2012.

[5] A. C. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, Chapter 7: Algebraic reconstruction algorithms, IEEE Press, pp.275, 1988.

[6] G. Wang and M. Jiang, "Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART)," J. X-Ray Sci. Technol., vol.12, pp.169, 2004.

[7] M. Lustig, D. L. Donoho and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," Magn. Reson. Med., vol.58, pp.1182, 2007.

[8] H. Yu and G. Wang, "Finite detector based projection model for high spatial resolution," J. X-Ray Sci. Technol., vol.20, pp.229, 2012.

[9] M. Jiang and G. Wang, "Convergence of the simultaneous algebraic reconstruction technique (SART)," IEEE Transactions on Imaging Processing, vol.12, pp.957, 2003.

[10] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," IEEE ISBI, pp.257, 2010.

[11] S. Ramani, P. Thevenaz and M. Unser, "Regularized Interpolation for Noisy Data," Proc. ISBI, pp.612, 2007.

[12] B. De Man, S. Basu, C. Naveen, B. Dunham, P. Edic, M. Maria, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower and E. Williams, "CATSIM: A new Computer Assisted Tomography SIMulation environment," Proc. SPIE, v.6510, doi:10.1117/12.710713, 2007.

[13] K. M. Hanson, "Detectability in computed tomographic images," Medical Physics, vol.6, pp.441, 1979.

Penalized Maximum Likelihood Iterative Reconstruction for Sub-wavelength Nano-scale Optical Computed Tomography (SNOCT) for Isolated Cell Imaging

Joyoni Dey[1], *Member IEEE*, Yong Wu[2], , Xundong Wu[2], Ming Xia[2], Zhongbo Yan[2], Enrico Stefani[2]and Ya-Hong Xie[2]

***Abstract*—Optical Computed (or Projection) Tomography is an exciting emerging field. Sharpe et al (*Science*, 2002), Fauver & Siebel et al (*Optical Express*, 2005), Miao et al, (*J. of Biomedical Optics*, 2009) used optical lenses to obtain approximately parallel projections (extended depths of fields) of a cellular object and used the Filtered Back Projection (FBP) algorithm for reconstructing 3D images of cells. They reported in their systems a resolution of 10-0.35 µm. Sun et al, (*J. of Applied Physics*, 2010) fabricated atomically sharp nano-tip tungsten field-emission arrays. Based on this prior work, as well as optical response studies of graphene (Wang et al, *ACS Nano, 2012*), Ya-Hong Xie at UCLA conceived of a sub-wavelength, nano-scale optical computed tomographic system (SNOCT) where thousands of light emitters (built with phosphor layers in between field-emitter (cathode) and graphene rod anodes), emitting green light at ~540nm wavelength. The object is placed close to the light source and light is detected by a CMOS photo-detector array. The aim is to reconstruct the total attenuation profile in isolated cells. The system has no high precision moving parts or optical lenses and is expected to yield sub-micron resolution. In this work we show one potential (cone-beam) geometry of the system with 49 source positions spaced $3\mu m$ apart with ~100 degree coverage. We built a cone-beam forward simulator and a penalized maximum likelihood algorithm to reconstruct example objects. The ROI is $10\mu m \times 10\mu m \times 10\mu m$. A uniform object and an object with an off-axis spherical attenuator yielded reconstruction results with overall 2.2% absolute error or less (from original object). The current system resolution is 0.5 microns isotropic (reconstruction speed being the primary limitation). Ongoing and future work involves a full forward simulator of the propagation through tissue using the P3 approximation of the radiation transport equation; speeding up the reconstruction to enable sub-micron-scale resolution reconstruction.**

I. Introduction

OPTICAL Projection or Computed Tomography (or OPT in short, [1]-[3] is an emerging modality which allows rapid 3D mapping of tissue distribution using multi-projection illumination, instead of obtaining data in "optical sections" used in other conventional microscopic systems, (confocal microscopy with de-convolution, and optical coherence tomography). Acquiring 3D data (transmission and fluorescence imaging) rapidly allows for basic biological temporal studies of growth of embryos, organelles, RNA distribution, protein expression, gene function, stimulus response of cells and so on [1] and has the potential to advance the field of basic biological research tremendously.

In [1]-[3] optical lenses were used to obtain near-parallel projection and thus FBP reconstruction was used. In [1] high resolution 3D volumes of both fluorescent and non-fluorescent biological specimen up to 15 mm thickness could be imaged with an expected resolution of 5-10 micron and 400 projection images were used for FBP reconstruction. In [2] cellular nuclei were imaged using 250 projections (extended depths of fields) from 0-180 deg and FBP reconstruction, with 0.9 micron of expected resolution. In [3] fluorescent images were also used and the expected resolution was 0.35 microns.

Sun et al [4] fabricated atomically sharp nano-tip tungsten field-emission arrays. Based on this methodology, as well as optical response studies of graphene [5], Xie conceived of a sub-wavelength, nano-scale optical computed tomographic system (SNOCT) where thousands of light emitters (built with phosphor layers in between field-emitter (cathode) and graphene rods (anode), emitting green light at ~540nm wavelength. The object is placed close to the light source and light is detected by a CMOS photo-detector array. The aim is to reconstruct the total attenuation in isolated cells. The system is expected to yield sub-micron resolution. The system has no high precision moving parts or optical lenses and is expected to be significantly less costly than a high-end microscope costing ~$100K-$1000K.

In this work we investigate a geometric configuration for SNOCT. There are multiple stationary source field emitters and the detector is a panel of photo-detector arrays (Figure 1a). The object is placed close to the source to obtain the desired magnification. The source field emitters may be distributed about 20nm apart. In between field emitter (cathodes) and anodes made of graphene (useful for near

Joyoni Dey is with University of Massachusetts Medical School, Worcester, MA. Contact email: joyoni.dey@umassmed.edu. Yong Wu, Xundong Wu, Ming Xia, Zhongbo Yan, Enrico Stefani and Ya-Hong Xie are at the University of California, Los Angeles.

optical transparency), a thin film of electroluminescent phosphor layer converts the kinetic energy of the field emitted electrons into photons. Each source could be turned ON/OFF. The projection of each source is obtained individually. The sources emit green-light (540nm wavelength) but other wavelengths (near-infrared) are possible. SNOCT may also be used in fluorescence mode.

II. Methods

A. System Geometry

Figure 1(b) shows a potential system geometry of SNOCT. The detector panel is 1cm (can be 1 to 2 cm) with photo-detectors that are 0.5micron apart. The source to detector distance is $A = 2mm = 2000\mu m$. The region of interest is $10\mu m \times 10\mu m \times 10\mu m$ and the object center is $H = 12\mu m$ from the central source. This would afford a magnification factor of $A/H = 2000/12 = 166.7$. Thus the smallest voxel size that would be mapped to a detector pixel of $\Delta x = 0.5\mu m$ is $\Delta v = 0.5 \times 12/2000 = 0.003\mu m = 3nm.$ Thus as far as the geometry is concerned at least we can expect to be able to reconstruct in tens of nanometers. The width of the region of interest $W = 10\mu m$ is mapped to $X = 1.67mm$. With our simplest geometry, resolving objects in the directions parallel to the detector (x-z direction here) might be easier than that in the source-detector direction (y-direction). In the y-direction (called axial here), the net attenuation presented by the object of $10\mu m$ thickness would be about 10^{-4}. The dynamic range of detector is 100dB, allowing detection of a 10^{-5} in amplitude (i.e. only about 10 times smaller than the total attenuation). One concern is that scatter may occupy some of the detector range. Total attenuation due to (back-) scatter and absorption is a parameter we expect to image (as per radiative-transport equation), but scattered photons detected from different directions is expected to be low because of the large detector distance from the object compared to the size of the object. Scatter appears not to pose challenges in [1]-[3] where a FBP reconstruction yielded good results. In ongoing/future work we will theoretically and experimentally investigate the effects of scatter. A 7x7 grid of emitters is chosen from the source panel. The sources are to be turned on one at a time. However, from diffraction limit considerations, a spacing of $3\mu m$ (or about 5.5 x the wavelength), is used to make sure that the projections from one emitter to next will yield adequately distinct information. The angular coverage is +/-49.5 degs in two directions. The 7x7 grid ensures no projection is truncated by the 1cm wide detector. One can add other detector modules horizontally or vertically in two sides to have more angular coverage. In this work we will investigate the simplest single-detector panel geometry to see if it can give accurate reconstruction using iterative reconstruction methods.

B. Forward Model and Reconstruction

Analytical forward model is developed using Siddon-like ray-tracing. We use non-subset version of the Penalized Maximum Likelihood (PML) reconstruction algorithm [6]-[7] for Computed Tomography which maximizes the likelihood function while also including a regularizing prior has better handing for truncation effects, noise handling. The PML algorithm uses breaks the likelihood function and the regularizing prior into separable parabolic surrogates and Newton's method is used to optimize the functional. The update equation for the attenuation μ at the jth voxel is given by

$$\mu_j^{n+1} = \mu_j^n + \frac{\dot{L}_j - \beta \sum_k w_{jk} \dot{\psi}(\mu_j^n - \mu_k^n)}{d_j^* + 2\beta \sum_k w_{jk} w_\psi (\mu_j^n - \mu_k^n)}$$

where ψ is a prior function in this case chosen to be the Huber prior; $w_\psi = \dot{\psi}(x)/x$ is an approximation of the second derivative of the function. β is a parameter which dictates the relative weight of the likelihood and the smoothing prior. If some strict conditions (e.g. monotonicity) the second term in the numerator is not used and to save CPU-time, the d_j curvature can be pre-computed [6]. We use the simplified version of the equation. The β is optimized for an example case of an uniform object.

III. Results

A. Projection (Forward Model)

The analytic forward model of the system is used to obtain 49 projections for different objects. The source intensity is taken as unity. The voxel values of objects are the order of magnitude for total-attenuation in tissue *per-micron*. Figure 2 shows all the 49 projections for the 2000x2000 detector (1cm x1cm at $0.5\mu m$) for a uniform object. We note that none of the projections are truncated.

B. Reconstruction

The PML algorithm is used for reconstruction. A 20x20x20 volume is reconstructed with $0.5\mu m = 500nm$ voxel size. This takes 28 sec/iteration. The reason for choosing this resolution is that reconstruction is as yet too slow for a smaller voxel size (e.g. 200x200x200 volume at $50nm$). A reconstructed uniform object is shown in Figure 3. Y-direction is along source to detector. Slices in x-z and x-y direction are shown. While small, the y-direction is slightly more inhomogeneous than other directions, as might be expected for limited angle reconstructions. However quantitatively, the bias (error) is small in all directions and marginally different in the y-direction from other directions. The β is broken into two parts $\beta = u_0 b_0$. For $u_0 = 0.1$, the parameter b_0 is varied in three orders of magnitude: 10^4,$10^3$and 10^2. The net absolute error or bias averaged across the entire object volume, leaving out a few bordering pixels in each direction is measured with iteration. After 120 iterations (nearly converged) the best performance is obtained around b_0=10^3. The bias at 120 iterations for this β is only 2.25%. The reconstruction is shown with the final parameter for an uniform object and object with an off-axis spherical attenuator. A point on the sphere is tracked to have a low bias of 1.5% after 120

iterations. Note a different β value maybe optimum for different source intensity. Also $u_0 = 0.1$ may be optimized.

IV. Conclusion And Future Work

With limited 99 degs coverage (in two directions), initial choice of geometry worked to reconstruct objects (uniform object and uniform object with a sphere) with under 2.2% bias. More angles may be obtained by using wider detector (2x2cm), and other orthogonal detector panels. In the future, the reconstruction is to be speeded up (expected by order of 40) by considering only the boundary of the ROI projected onto the detector and ordered subsets. Since the forward projection is the primary speed bottleneck, significant speed ups can be expected upon parallelizing (e.g. all the projections in parallel CPUs). After these speed-ups a nano-resolution (~50nm) volume reconstruction can be expected to be about an hour.

A fuller forward model is also being investigated where the full effects of scatter is investigated using the radiative transport equation, where, apart from the total attenuation, a scatter phase function is assumed to account for the detected photons scattered by tissue into different directions. This will be compared with Monte Carlo simulation and experiments.

References

[1]J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sorenen, R. Baldock and D. Davidson, "Optical Projection Tomography as a Tool for 3D Microscopy and Gene Expression Studies", *Science*, 2002 Vol. 296, No. 5567, Apr. 19, 2002

[2]M. Fauver, E. Siebel, J.R. Rahn, M.G. Meyer, F. W. Patten, T. Neumann, A.C. Nelson, "Three-dimensional imaging of single isolated cell nuclei using optical projection tomography", *Optical Express*, vol. 13, May 2005

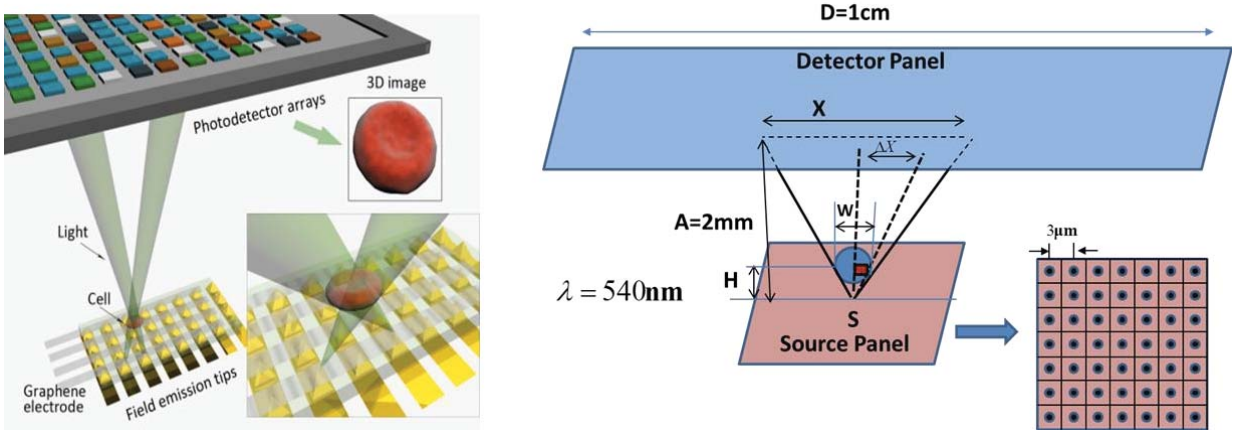
[3]Q. Miao, J. R. Rahn, A.Tourovskaia, M. G. Meyer, T. Neumann, A. C. Nelson, E. J. Seibel "Dual-modal three-dimensional imaging of single cells with isometric high resolution using an optical projection tomography microscope", *J. of Biomedical Optics*, vol. 14, no. 6, Nov/Dec 2009

[4] K. Sun,J. Y. Lee, B. Li, W. Liu, C. Miao, Y-H. Xie, X Wei, and T. P. Russell, "Fabrication and field emission study of atomically sharp high-density tungsten nanotip arrays", *J. Appl. Phys*. vol. 108, 036102, 2010

[5] P.Wang, W. Zhang, O. Liang, M. Pantoja, J. Katzer, T. Schroeder, and Y-H. Xie, "Giant Optical Response from Graphene_Plasmonic System" *ACS Nano,* vol. 6, no. 7, pp. 6244–6249, 2012

[6] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography", *Phys. Med. Biol*, vol. 44, pp. 2835-2851, 1999

[7] J. A. Fessler, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis,* M. Sonka and J. Michael Fitzpatrick, editor. SPIE, Bellingham, pp. 1-70, 2000.



(a) (b)

Figure 1 (a) Schematic Diagram of SNOCT. (b) Schematic details of the geometry and source distribution chosen.

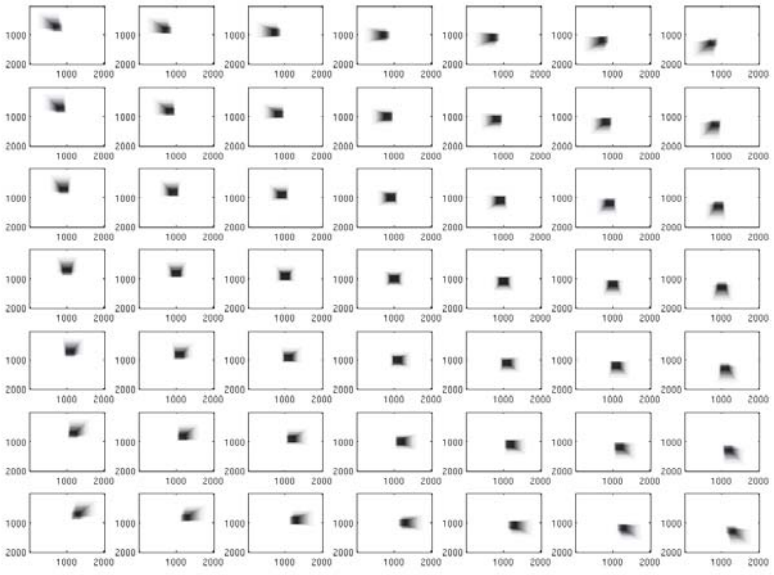
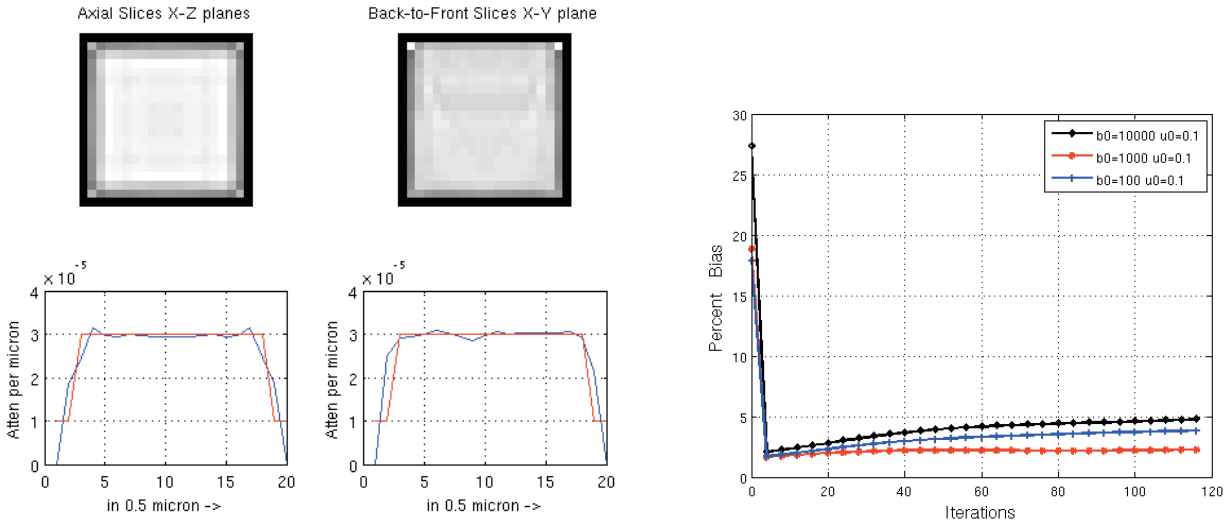
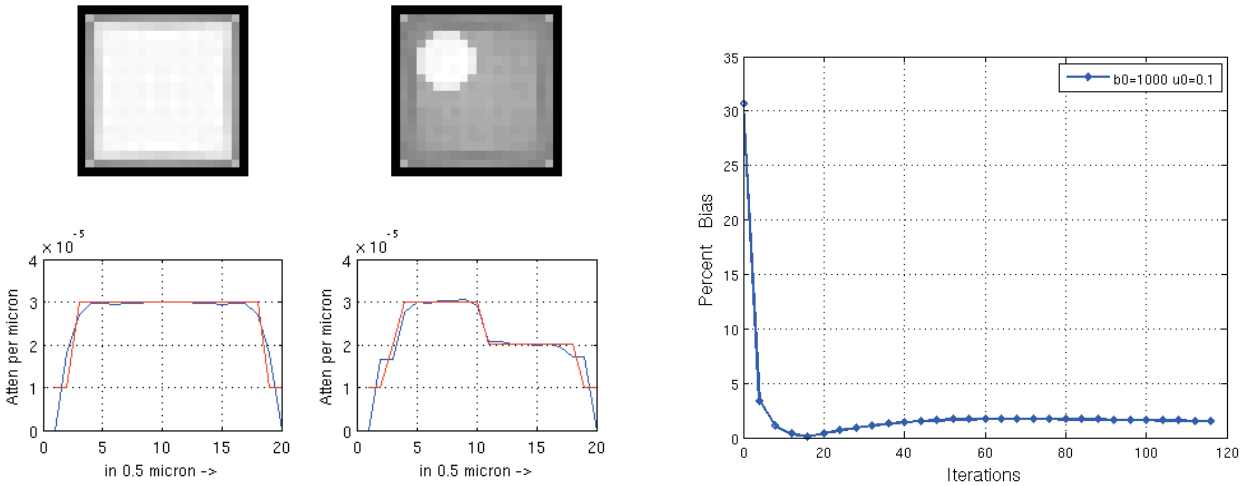


Figure 2. The 49 different projections



(a) (b)

Figure 3(a) X-Z plane slice (top-left) and a profile (bottom-left) for an uniform object. X-Y plane slice (top-right) and a profile (bottom-right) for same object. The y-axis is from source to detector. (b) Percent-error or Percent bias versus iteration plots for 3 different β.



(a) (b)

Figure 4(a) X-Z plane slice (top-left) and a profile (bottom-left) for an uniform object. X-Z plane slice (top-right) and a profile (bottom-right) for a uniform object with a off-axis sphere. The y-axis is from source to detector. (b) Percent bias versus iteration plot for a point on the sphere.

Image Grid Invariant Regularization for Iterative Reconstruction

Zhou Yu, Charles A. Bouman, Jean-Baptiste Thibault, and Ken D. Sauer

***Abstract*— Regularization methods have been successfully applied to various reconstruction and denoising problems. In these problems, one needs to choose regularization parameters that properly balance resolution and noise. These parameters need to be adjusted when the image grid pattern changes. In this paper, we present a theory on regularization design derived from a continuous image model so that the regularization method is invariant to image grid pattern. We can use this theory to compute regularization parameters for various image grid patterns. To illustrate the idea, we applied this theory to regularization design on a rectangular image grid.**

***Index Terms*— Iterative reconstruction, regularization method, image model, multi-grid**

I. Introduction

Iterative reconstruction (IR) methods have been widely applied to solve various reconstruction and denoising problems in medical imaging, security CT, electron-microscopy, etc [1]–[3]. IR methods typically solve the reconstruction problem by formulating mathematical models of the physics and statistics of the imaging process and the image itself.

One important aspect of the modeling is to find a discrete model of the image and physical processes which are generally speaking continuous in nature. In a p-dimensional reconstruction problem, we typically model the image object as a continuous function $f(r) : \Re^p \mapsto \Re$, where r is the vector representing spatial location. The input to a reconstruction problem is a set of discrete measurements denoted by a vector y. The output of the reconstruction is a discrete image array denoted by vector x. To define a discrete representation of the image, one can define x to be the samples of f,

$$x_i = f(r_i) \tag{1}$$

where i is the pixel index and r_i are typically chosen to fall on a periodic grid.

The imaging process can be modeled as a mapping from f to y, that is, $y = F(f)$. For example, in 2D parallel beam CT reconstruction, F is the Radon transform. Once we define a discrete representation of f, we can also derive a discrete forward model, $y \approx \tilde{F}(x)$, mapping from x to y. In the CT reconstruction problem, for instances, various discrete forward models have been proposed [4]–[7]. Once the models are built, one can formulate a cost function to find the solution that best fits the model. In general, the image is reconstructed by minimizing a cost function such as,

$$\hat{x} = \arg\min \left\{ G(\tilde{F}(x), y) + U(x) \right\}, \tag{2}$$

in which $G(\tilde{F}(x), y)$ is the data mismatch term that penalizes the differences between the image x and the measurement y according to the forward model $\tilde{F}$ and a statistical model, and $U(x)$ is the regularization function that penalizes the roughness in the image [8]–[10]. $U(x)$ is typically derived from a prior model of the image, such as the Markov random field (MRF) model [11]. The regularization term $U(x)$ plays a very important role in defining the image quality [12].

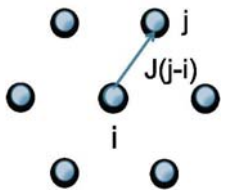


Fig. 1. This figure shows an example neighborhood on a non-rectangular periodic 2D image grid.

In many imaging problems, the image grid might vary from case to case. For example, when reconstructing images on rectangular grids, the sampling rate along each direction might vary depending on the desired resolution. In other cases, the images might not necessarily fall on rectangular grids. Therefore we would like to design a cost function that is invariant to the image grid. In equation (2), the discrete forward model $\tilde{F}$ naturally takes the pixel size and location into consideration. In equation (28) of [13], Oh et.al proposed a scale invariant design of $U(x)$, assuming the sampling pattern (in this case, the pixel's aspect ratio) remains the same. In this paper, we propose a general theory to design $U(x)$ to be image grid invariant. To do this, we first approximate $U(x)$ to be an integral of f. In order to be image grid invariant, this integral must not depend on the pixel location or the choice of the neighborhood. Based on this condition, we can derive a set of sufficient conditions to guide the design of $U(x)$. We found that these conditions can be satisfied in the case of quadratic regularization or when $U(x)$ has a special form. To provide an example of the theory, we describe a design for a commonly used 3D rectangular grid in section III. Finally, we apply the design to the 3D CT reconstruction problem as an example.

II. Theory

Let us first introduce notation to describe a periodic image grid. We use $i = (i_1, i_2, \ldots, i_p)$ and $j = (j_1, j_2, \ldots, j_p)$ to denote p-dimensional discrete indexes; and s to denote the coordinates of a pixel. Let S be a lattice of pixels at locations

Zhou Yu and Jean-Baptiste Thibault are with GE Healthcare, 3000 N Grandview Bvd, W-1180, Waukesha, WI 53188.

Charles Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-0501.

Ken Sauer is with the Department of Electrical Engineering, 275 Fitzpatrick, University of Notre Dame, Notre Dame, IN 46556-5637.

$s_i \in \Re^p$ taking on the values $f(s_i)$. Furthermore, let $\mathcal{P} \subset S^2$ be a neighborhood system on S, where $\mathcal{P}$ consists of all voxel pairs $\{i, j\}$ where i is a neighbor of j. We assume that the lattice S has a periodic structure, so the neighbors of $i \in S$ are $j = i+k$ where $k \in W$, and W is a set of neighboring pixels' index displacements. We also assume that W is symmetric, so that if $k \in W$ then $-k \in W$. We define J as a p by p transform matrix that computes the displacement of two neighboring pixels from k,

$$s_j - s_i = Jk. \tag{3}$$

In this notation, the neighborhood is defined by the set W describing the selection of the neighboring pixels and the matrix J describing the displacement between each neighboring pixel pair. Fig. 1 shows an exemplary neighborhood on a non-rectangular periodic 2D image grid, in which each pixel has 6 neighbors.

Let us also assume the prior distribution of x as a Markov random field (MRF), and we will discuss a general form later. In the MRF case, $U(x)$ is of the form,

$$U(x) = \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} b_{j-i} \rho(x_j - x_i) \tag{4}$$

where $\rho(\cdot)$ is the potential function. Our objective is to design α and b_{j-i} so that $U(x)$ is invariant to the image grid.

We first derive $U(x)$ as a discrete approximation to an integral of f. To do this, we use the finite difference in the image to approximate the local directional gradient of f, that is,

$$x_j - x_i \approx \nabla f(s_i)^t J(j-i). \tag{5}$$

Using this approximation, we can rewrite $U(x)$ as,

$$\begin{aligned} U(x) &= \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} b_{j-i} \rho(x_j - x_i) \\ &= \frac{1}{2\alpha} \sum_{i \in S} \sum_{k \in W} b_k \rho(x_{i+k} - x_i) \\ &\approx \frac{1}{2\alpha} \sum_{i \in S} \sum_{k \in W} b_k \rho(\nabla f(s_i)^t Jk) \\ &= \frac{1}{2\alpha|J|} \sum_{k \in W} b_k \sum_{i \in S} \rho(\nabla f(s_i)^t Jk)|J| \\ &\approx \frac{1}{2\alpha|J|} \sum_{k \in W} b_k \int_{\Re^p} \rho(\nabla f(s)^t Jk) ds, \end{aligned} \tag{6}$$

where $|J|$ is the determinant of matrix J.

Notice that, in equation (6), the right hand side is a summation of integrals, in which each integral is based on the gradient along the direction of a specific neighbor, k. Since the right hand side approximation still depends on the choice of J and W, it is not image grid invariant in general. However, in some special cases, it is possible to use b_k to compensate for the directional change in the neighborhood. In the following, we derive a sufficient condition that b_k needs to satisfy to yield an image grid invariant regularization in two cases. In the first case, ρ must be a quadratic function. In the second case, ρ can be of a general form, however we need to introduce a minor modification to the form of $U(x)$.

A. Quadratic Regularization

Assume $\rho(\Delta) = \Delta^2$, then

$$\begin{aligned} U(x) &\approx \frac{1}{2\alpha|J|} \int_{\Re^p} \sum_{k \in W} b_k \nabla f(s)^t Jkk^t J^t \nabla f(s) ds \\ &= \int_{\Re^p} ||\nabla f(s)||_H^2 ds \end{aligned} \tag{7}$$

where,

$$H = \frac{1}{2\alpha|J|} \sum_{k \in W} b_k J(kk^t) J^t \tag{8}$$

To use equation (8) to design $U(x)$, we first need to choose a desired H matrix, such as an identity matrix. We then choose the image grid and neighborhood, i.e. J and W. Finally, we find b_k so that the equation (8) holds. Therefore, equation (8) gives a sufficient condition for image grid invariant regularization design.

B. General Potential Function

In many imaging applications, $\rho(x)$ is designed to suppress noise while preserving spatial resolution. To apply the theory to general potential functions, we would like to propose a different form of $U(x)$, where

$$U(x) = \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} \rho(b_{i-j}^2 (x_j - x_i)^2) \tag{9}$$

In this form, we sum over the squared difference between each neighboring voxel pairs first, then apply the penalty function $\rho(\cdot)$. Similar to the derivation of equation (6) and (7) , $U(x)$ can be shown to approximate the following integral,

$$U(x) \approx \frac{1}{2\alpha|J|} \int_{\Re^p} \rho(||\nabla f(s)||_H^2) ds, \tag{10}$$

where H is given by equation (8)

III. Application to 3D Rectangular Grid

In this section, we would like to provide an example design for a 3D rectangular image grid, with voxels spacing $\Delta_x = \Delta_y$, and Δ_z, where Δ_x and Δ_y are voxel sizes along x, y and z axis respectively.

We assume W to be a 3 by 3 by 3 cubic neighborhood, therefore, $W = \{-1, 0, 1\}^3$. In this case,

$$J = \begin{bmatrix} \Delta_x & 0 & 0 \\ 0 & \Delta_y & 0 \\ 0 & 0 & \Delta_z \end{bmatrix}$$

The geometry of the voxel neighborhood is illustrated in Fig. 2, in which θ is the angle between $k = (1, 0, 1)^t$ and the x axis as shown in (a), and γ is the angle between $k = (1, 1, 1)$ and the $x - y$ plane as shown in (b). Thus, $\cos\theta = \frac{1}{\sqrt{1+\beta^2}}$ and $\cos\gamma = \frac{\sqrt{2}}{\sqrt{2+\beta^2}}$, where $\beta = \frac{\Delta_z}{\Delta_x}$ is the aspect ratio of the voxel in this particular case where $\Delta_x = \Delta_y$. When the aspect ratio changes, the direction between two neighbor voxels, Jk, also changes. Without proper compensation using b_k, the regularization might be stronger in one direction relative to

the other. Let us assume we would like the regularization to be isotropic. Therefore, we choose $H = I$. In the following we will derive a solution for b_k so that equation (8) holds in this case.

Let $R_k = \frac{(Jk)^t Jk}{||Jk||^2}$, so that we can rewrite equation (7) as

$$\frac{1}{2\alpha|J|}\sum_{k\in W} b_k||Jk||^2 R_k = H \tag{11}$$

A sufficient condition for equation (11) to hold is $\alpha = \frac{1}{2|J|}$, $b_k = \frac{w_k}{||J_k||^2}$, where w_k satisfy,

$$\sum_{k\in W} w_k R_k = H \tag{12}$$

Notice that, $||J_k||$ is the distance between the two-voxel pair $(i, i+k)$. Therefore, in this case, b_k is inversely proportional to the squared distance, and w_k is computed to compensate for the different effective regularization strength along the direction of the neighbors defined by the image grid. We can derive the R_k matrices for each direction as follows:

$$R_{\pm1,0,0} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, R_{0,\pm1,0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$R_{0,0,\pm1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}; R_{\pm1,\pm1,0} = \begin{bmatrix} \frac{1}{2} & \pm\frac{1}{2} & 0 \\ \pm\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$R_{\pm1,0,\pm1} = \begin{bmatrix} \cos^2\theta & 0 & \pm\sin\theta\cos\theta \\ 0 & 0 & 0 \\ \pm\sin\theta\cos\theta & 0 & \sin^2\theta \end{bmatrix};$$

$$R_{0,\pm1,\pm1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos^2\theta & \pm\sin\theta\cos\theta \\ 0 & \pm\sin\theta\cos\theta & \sin^2\theta \end{bmatrix};$$

$$R_{\pm1,\pm1,\pm1} = \begin{bmatrix} \frac{1}{2}\cos^2\gamma & \pm\frac{1}{2}\cos^2\gamma & \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma \\ \pm\frac{1}{2}\cos^2\gamma & \frac{1}{2}\cos^2\gamma & \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma \\ \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma & \pm\frac{1}{\sqrt{2}}\sin\gamma\cos\gamma & \sin^2\gamma \end{bmatrix} \tag{13}$$

In general the solution to equation (12) might not be unique. In the following we will derive a solution for w_k when $H = I$. Instead of solving for 26 unknown w_k coefficients, we propose to apply a constraint on the solution such that neighbor voxel pairs of symmetric directions will have the same w_k values. This allows us to reduce the problem to 7 unknown variables. We assign the weight w_x to the direction $k = (\pm1, 0, 0)$ and w_y to $k = (0, \pm1, 0)$. Next, we assign w_{xy} to the set of directions $k = (\pm1, \pm1, 0)$, so that

$$\sum_{k_1=\pm1,k_2=\pm1,k_3=0} w_{xy} R_{k_1,k_2,k_3} = 4w_{xy}\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Notice that the matrices sum up to be a diagonal matrix. Similarly, we assign w_z to directions $k = (0, 0, \pm1)$, w_{xz} to directions $k = (\pm1, 0, \pm1)$, w_{yz} to directions $k = (0, \pm1, \pm1)$, and w_{xyz} to the directions $k = (\pm1, \pm1, \pm1)$

Substituting (13) into (12), we can verify that by setting the coefficients in the symmetric way, the non-diagonal entries will all cancel out. Therefore, we only need to consider the constraints of the three diagonal entries, which give us the following equations

$$\begin{cases} 2w_x + 4\times\frac{1}{2}w_{xy} + 4w_{xz}\cos^2\theta + 8\times\frac{1}{2}w_{xyz}\cos^2\gamma = 1 \\ 2w_y + 4\times\frac{1}{2}w_{xy} + 4w_{yz}\cos^2\theta + 8\times\frac{1}{2}w_{xyz}\cos^2\gamma = 1 \\ 2w_z + 4w_{xz}\sin^2\theta + 4w_{yz}\sin^2\theta + 8w_{xyz}\sin^2\gamma = 1 \end{cases} \tag{14}$$

The above equations still do not have a unique solution. Here we propose to formulate a cost function that minimizes the total energy of the w_k coefficients subject to (14) and a non-negativity constraint.

$$\begin{aligned} w^* &= \arg\min_{w_k\geq0}\sum_{k\in W} w_k^2 \\ &= \arg\min_{w_k\geq0}\{2w_x^2 + 2w_y^2 + 2w_z^2 \\ &\quad 4w_{xy}^2 + 4w_{xz}^2 + 4w_{yz}^2 + 8w_{xyz}^2\} \end{aligned} \tag{15}$$

The above constrained optimization problem yields the following analytical solution

$$\begin{bmatrix} w_x \\ w_y \\ w_{xy} \\ w_z \\ w_{xz} \\ w_{yz} \\ w_{xyz} \end{bmatrix} = \begin{bmatrix} 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/6\,\frac{12\beta^2+2\beta^4+12+\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{(1+\beta^2)(5\beta^4+4+16\beta^2+2\beta^6)}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/6\,\frac{(2+\beta^2)(5\beta^4+1+2\beta^2+\beta^6)}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \end{bmatrix}, \tag{16}$$

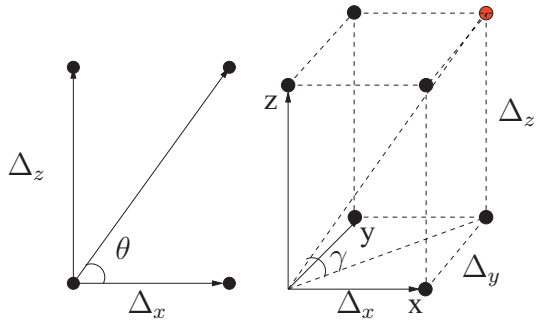
where $\beta = \frac{\Delta_z}{\Delta_x}$. Fig. 3 shows w_k as a function of β.

IV. Experimental Results

In this section, we apply the proposed regularization model to 3D CT reconstruction problems. In computed tomography, the resolution and noise properties are mainly determined by system geometry and scan techniques. Therefore, this paper does not focus on achieving specific image quality properties such as uniform or isotropic resolution in the general case for any scanning geometry. Instead, we would like to demonstrate that with the proposed model, image quality is less sensitive to voxel size changes compared to the baseline methods.

The methods in comparison are the proposed regularization model, and two baseline methods. In the first baseline method, we choose $\alpha = \frac{1}{2|J|}$ and $b_k = \frac{1}{||Jk||^2}$, that is, we set w_k to be a constant in all directions. In the second baseline method, we choose $b_k = \frac{1}{||Jk||^2}$ and α to be a constant for all voxel sizes. In both baseline methods, we adjust the overall regularization strength to match the proposed model at $dx = dy = dz = 0.625mm$. At this point, all three models yield identical regularization parameters.

To measure image quality, we simulated a 3D digital phantom containing an array of high contrast point sources as shown in Fig. 4. We test the proposed and baseline algorithms with various voxel sizes $dx = dy$ at fixed $dz = 0.625$mm. In Fig. 5, we measure the 50% MTF (lp/cm) in $x-y$ plane, full width half maxium (FWHM) in mm along the z axis and noise standard deviation (HU) in the uniform portion of the phantom. Considering the variations of these image quality metrics when the voxel size changes, we notice that the proposed method is less sensitive to the voxel size change compared to the baseline methods. Furthermore, the baseline 1 is also less sensitive compared to the baseline 2, since it is closer to the proposed model.



(a) (b)

Fig. 2. This figure illustrates the geometry of the voxel neighborhood on a rectangular grid. In (a), we show the voxels in the $x-z$ plane. The angle between the diagonal voxel and the x axis is denoted as θ. In (b), we show the voxels in the first octant, in which γ is the angle between the corner voxel shown in red a

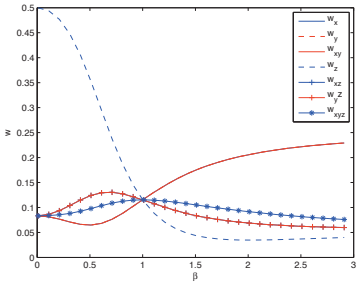


Fig. 3. This figure shows w_k as a function of voxel aspect ratio parameter β

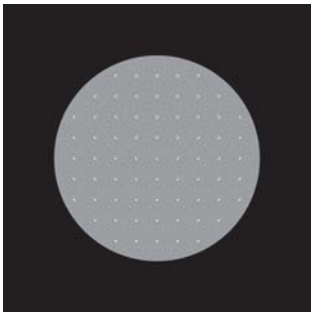


Fig. 4. The figure shows the digital phantom used in the experiment.

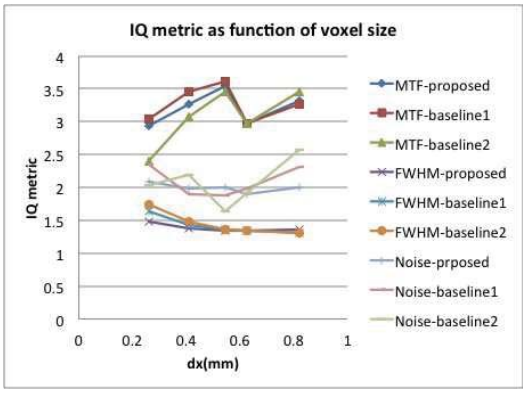


Fig. 5. The figure shows the IQ metrics when $dx = dy$ changes at fixed dz. The proposed method appears less sensitive to such changes compared to the baseline methods.

V. Discussion

The theory proposed in this paper can be useful in many reconstruction as well as denoising problems. In our example, we change the pixel size independently of slice thickness, which yields a non-isotropic voxel. Without modeling non-isotropic sampling in $U(x)$, one could end up with over-regularized results along certain directions while other directions are under-regularized. In 4D reconstructions (3D + temporal), the voxel size and temporal sampling rate can also be adjusted independently. The proposed theory can be used to yield a consistent design balancing spatial and temporal resolution. Although our derivation assumes a periodic lattice, the theory can also be extended to non-periodic lattices where b_k needs to be computed for each location.

References

[1] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.

[2] E. Ü. Mumcuoğlu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[3] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[4] P. Joesph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Trans. on Medical Imaging*, vol. 1, pp. 192–196, 1982.

[5] R. Lewitt, "Multidimensional digital image representations using generalized kaiser-bessel window functions," *J. Opt. Soc. Am. A*, vol. 7(10), pp. 1834–1846, October 1990.

[6] B. DeMan and S. Basu, "Distance-driven projection and backprojection in three-dimensions," *Physics in Medicine and Biology*, vol. 49, pp. 2463–2475, 2004.

[7] A.Ziegler, Th.Köhler, and R.Proksa, "Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry," *Med. Phys.*, vol. 33, no. 12, pp. 4653–4663, 2006.

[8] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, March 1992.

[9] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[10] V. Panin, G. Zeng, and G. Gullberg, "Total variation regulated EM algorithm," *IEEE Trans. on Nuclear Science*, vol. 46, no. 6, pp. 2202–2210, December 1999.

[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, pp. 721–741, November 1984.

[12] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[13] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 125–140, January 2005.

Ultrafast Fully Analytical Iterative Model-based Statistical 3D Reconstruction Algorithm

Robert Cierniak and Michal Knas

***Abstract*—The orginally formulated by us a 3D reconstruction algorithm for spiral cone-beam x-ray tomography is presented. This appraoch to the 3D reconstruction problem is based on the fully analytical formulation of the reconstruction problem. Additinally, statistical considerations involved with the form of noise present in x-ray measurement system at the formulation of this problem are taken into account. This conception significantly improves quality of the obtained after reconstruction images and decreases the complexity of the reconstruction problem in comparison with other approaches. The carried out computer simulations shown that presented here reconstruction algorithm outperforms conventional analytical methods in obtained image quality. The GPU realization of our algorithm shows that this algorithm can be fully applicable for commercial use in the sense of the obtained image quality and the time of reconstruction process.**

***Index Terms*—Computed tomography, 3D reconstruction algorithm, image reconstruction from projections problem.**

I. Introduction

Nowadays, statistical reconstruction methods are being actively developed because they allow us to improve the resolution of reconstructed images and/or decrease x-ray intensity while maintaining the quality of the CT images obtained. This is because the signals are adapted to the specific statistics of a given technique and consequently the algorithms can yield a reduction in radiation dose during human body examination [1]. It is worth to note that there are several commercially developed and introduced statistical iterative reconstruction algorithms and iterative image based denoising algorithms, e.g. Adaptive Statistical Iterative Reconstruction (ASIR), Iterative Reconstruction in Image Space (ARIS), Adaptive Iterative Dose Reduction (AIDR) or iDose algorithms.

Generally, one can say that all most significant existing reconstruction algorithms belong to two basic approaches, taking into account the methodology of the used in them signal processing concepts: these called the analytical methods, and those assigned to strategy called the algebraic reconstruction technique. We can suppose that the implementation of the ART in the historical first CT apparatus was caused for lack of alternative at that time. Next generation of CT systems used only reconstruction algorithms based on analytical methodology. The main reason of this situation was the huge sizes of matrices appeared in the algebraic reconstruction problem, and caused by this the calculation complexity of reconstruction method based on this methodology. The analytical methodology simplifies drastically the number of neccessary calculation and in this way is more appealing. The algebraic approach was taken firstly into consideration (see e.g. [2], [3]) for design of the statistical reconstruction algorithms because it allows for accurate modelling of the statistics of projection data and it helps to avoid most of distortion caused by them. Presented in above publications reconstruction idea is based on the maximum *a posteriori* probability (MAP) estimation approach. That application of the algebraic reconstruction technique has some significant technical difficulties at practical realization, namely: in the case of algorithms for 3D spiral cone-beam scanners, it is complicated to establish the coefficients of forward model for ART at this geometry of scanner [3], [5], [4]; this methodology forces simultaneously calculations for all voxels in range of reconstructed 3D image what makes the reconstruction problem extremely complex.

We could avoid the mentioned above difficulties connected with using of ART methodology using an analytical strategy of the reconstructed image processing. In previous papers we showed how to formulate the analytical reconstruction problem consistent with the ML methodology for parallel geometry of scanner [6]–[8], for fan-beams [9], and finally we proposed the scheme of reconstruction method for the spiral cone-beam scanner [10]. Our statistical 3D reconstruction algorithm has some serious adventages compared with algebraic methodology. Firstly, we establish certain coefficients in our method, but it is performed in much more easer way than in comparative method. Secendly, we can perform reconstruction process only in one plane in 2D space, what simplify the problem. The reconstruction process can be performed for every cross-section image separately. After this, it is possible to reconstruct whole 3D volume image from set of the reconstructed before 2D images. Additionally, we use during reconstruction process the FFT what significantly decreases the time of the reconstruction process. Moreover, we propose such modification of the imposed loss function in the applied ML estimation that it is unnecessary to use any additional regularization term (*a priori* term). Thanks to this modification we prevent any unstabilities in reconstruction process and we simplify the loss function adapted to statistics of projection data involved in the x-ray computed tomography.

II. 3D Reconstruction Algorithm for Spiral Cone-beam Scanner

We have formulated orginally the 2D analytical approximate reconstruction problem for parallel geometry of scanner (see e.g. [7], [8]). Presented there idea can be incorporated into the Feldkamp-type reconstruction methodology for design of 3D reconstruction algorithm for the spiral cone-beam geometry

Corresponding author: Robert Cierniak, Institute of Computational Intelligence, Czestochowa University of Technology, Armii Krajowej 36, 42-200 Czestochowa, Poland, e-mail: cierniak@kik.pcz.czest.pl

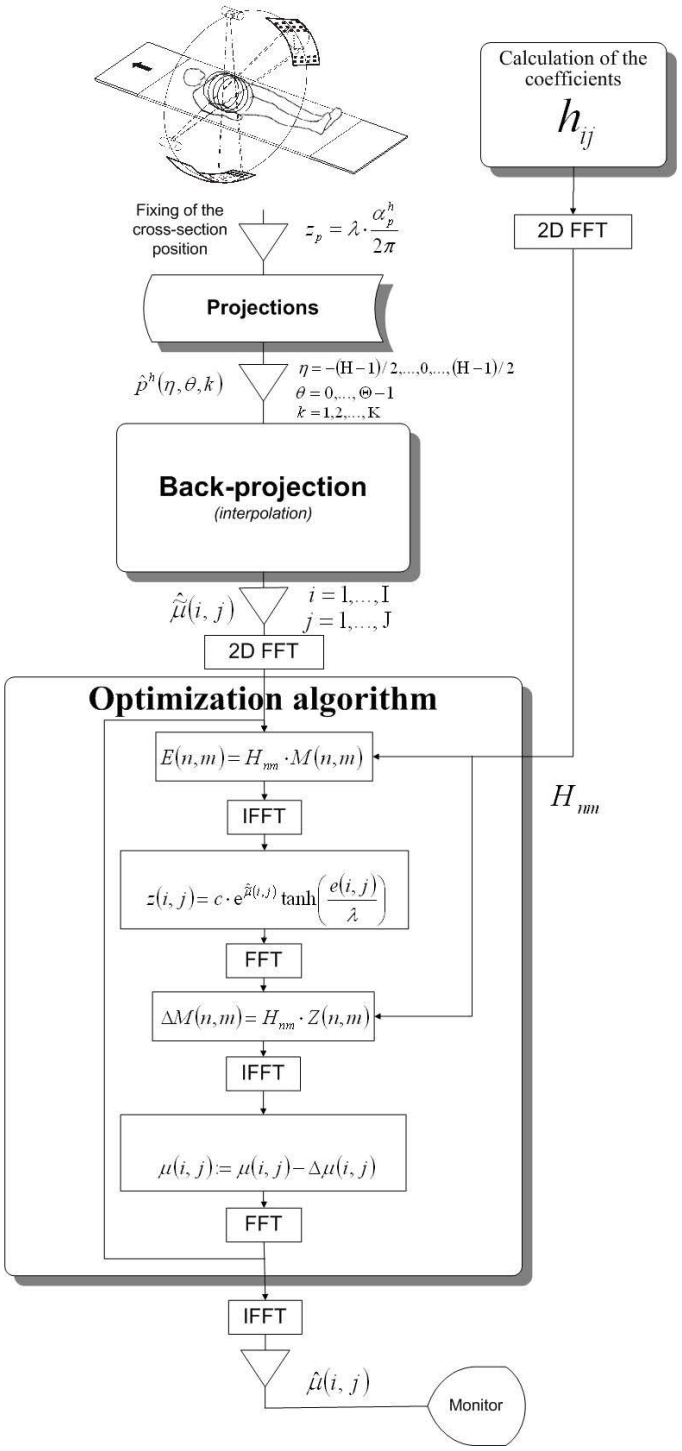


Fig. 1. An image reconstruction algorithm for the cone-beam geometry scanner

of the scanners. We will show in the following subsections how can adapt our orginal idea to the 3D reconstruction problem. Additionally, we will describe the statistical nature of our method, what lead to the design of the 3D iterative model-based reconstruction algorithm. The general scheme of proposed by us reconstruction procedure is depicted in Fig. 1.

Presented here iterative statistical reconstruction algorithm uses projections obtained in an spiral cone-beam scanner. After selecting a position z_p on the z-axis of the reconstructed cross-section of a body, the reconstruction of image at this position is performed.

In a real spiral cone-beam projection system, the reconstruction algorithm can only make use of projections obtained at certain angles and measured only at particular points on the partial cylindrical-shaped screen. In our case, the beam of x-rays reaches the individual detector rows $k = 1, 2, \ldots, \mathrm{K}$, where K is a number of rows placed on the screen. In every row, selected rays strike the detectors, each of which is indexed by the variable $\eta = -(\mathrm{H}-1)/2, \ldots, 0, \ldots, (\mathrm{H}-1)/2$, where H is an odd number of detectors in each channel of the array. Detectors are placed on the screen separated by a distance Δ_k in each row, and by an angular distance Δ_η in each channel. Additionally, only a limited number of projections are performed, each of which is indexed by the variable $\theta = 0, \ldots, \Theta - 1$, where $\Theta - 1$ is the total number of projections made during the examination. Every projection is carried out after rotation by Δ_θ. We can sum up these practical conditions by saying that the reconstruction algorithm has available to it the projection values $p^h\left(\beta_\eta, \alpha_\theta^h, \dot{z}_k\right)$, in the ranges: $\eta = -(\mathrm{H}-1)/2, \ldots, 0, \ldots, (\mathrm{H}-1)/2$; $\theta = 0, \ldots, \Theta - 1$; $k = 1, 2, \ldots, \mathrm{K}$.

Several following after the acquisition of projections process operations are the same as in the case of the reconstruction algorithm designed for parallel-beam scanner. The only defference between 2D and 3D approaches lies in performance of the back-projejection operation, and this operation is presented as the next one.

A. Back-projection Operation

It is at this stage of the process that the three-dimensional back-projection is performed. Every point in the coordinate space is given a value equal to the sum of all the projection values from all the rays passing through the point. For the projections $\dot{p}^h\left(\beta_{ij}, \theta, \dot{z}_{ij}\right)$ made at angle θ; $\theta = 0, \ldots, \Theta - 1$, the operation can be written:

$$\tilde{\mu}\left(i, j, z_p\right) \cong \Delta_\alpha^h \cdot \sum_\theta \dot{p}^h\left(\beta_{ij}, \theta, \dot{z}_{ij}\right), \tag{1}$$

Often there is no ray that actually passes exactly through a particular discrete point in space, or voxel, (i, j, z_p) and so there is no projection value available to the reconstruction algorithm. This means that the missing projection value has to be interpolated based on measurements $p^h\left(\eta, \theta, k\right)$, using for example the technique of bilinear interpolation:

$$\dot{p}^h\left(\beta_{ij}, \theta, \dot{z}_{ij}\right) = \sum_{n=1}^{4} c_{n,\theta,ij} p_{n,\theta,ij}^h \tag{2}$$

where the coefficients of interpolation $c_{n,\theta,ij}$ can be established in the following way:

$$c_{n,\theta,ij} = \left(1 - \frac{|\Delta\beta|}{\Delta_\eta}\right) \cdot \left(1 - \frac{|\Delta\dot{z}|}{\Delta_{\dot{z}_k}}\right), \tag{3}$$

where the quantities β_{ij} and $\dot{z}_{ij}$ represent the coordinates of the discrete point (i, j, z_p) expressed as parameters of the

projection carried out at the angle θ. It can easily be shown that $\dot{z}_{ij}$ is calculated as follows:

$$\dot{z}_{ij}(\theta) = \frac{\mathrm{R}_f \cdot (z_p - z_0(\theta))}{\mathrm{R}_f - u_{ij}} \tag{4}$$

where R_f is the radius of the circle described by the focus of the tube, and

$$z_0(\theta) = \lambda \cdot \frac{\theta \Delta_\alpha^h}{2\pi} \tag{5}$$

and

$$u_{ij} = i\Delta_{xy} \cdot \sin\left(\theta\Delta_\alpha^h\right) - j\Delta_{xy} \cdot \cos\left(\theta\Delta_\alpha^h\right), \tag{6}$$

where: Δ_{xy} is the interval between individual points on the reconstructed image.

The value of β_{ij} is easily determined from the following formula:

$$\beta_{ij}(\theta) = \arcsin\left(\frac{s_{ij}}{\sqrt{\mathrm{R}_f^2 + \dot{z}_{ij}} - u_{ij}}\right), \tag{7}$$

where

$$s_{ij} = i\Delta_{xy} \cdot \cos\left(\theta\Delta_\alpha^h\right) + j\Delta_{xy} \cdot \sin\left(\theta\Delta_\alpha^h\right), \tag{8}$$

and u_{ij} has already been determined in equation (6).

It is easy now to explain how are chosen these four projections $p^h_{n,\theta,ij}$ used in bilinear interpolation (2): they are determined as the nearest neighbours of the projection $\dot{p}^h(\beta_{ij}, \theta, \dot{z}_{ij})$, taking into account the parameters $\beta_{ij}(\theta)$ and $\dot{z}_{ij}(\theta)$ (equations (7) and (4), respectively).

B. Iterative Reconstruction Procedure

The main goal of the presented here iterative statistical procedure is to reconstruct image according to the maximum likelihood (ML) methodology of estimation of expected value of image for certain pixels, holding the analytical scheme of image processing in given reconstruction algorithm. After deep statistical analysis, we propose the following form of objective to be optimized during this iterative reconstruction procedure:

$$\mu^*_{\min} = \arg\min_{\mu^*}\left(\frac{1}{2}\sum_{i=1}^{\mathrm{I}}\sum_{j=1}^{\mathrm{J}}\frac{1}{\sigma^2_\Sigma(i,j)} f(e(i,j))\right), \tag{9}$$

where:

$$e(i,j) = \sum_{\bar{i}}\sum_{\bar{j}} \mu^*(\bar{i},\bar{j}) \cdot h_{\Delta i,\Delta j} - \tilde{\mu}(i,j), \tag{10}$$

and (what can be shown)

$$h_{\Delta i,\Delta j} = h_{i,j} = \Delta_\alpha \cdot \sum_{\psi=0}^{\Psi-1} i\bar{n}t\left(i\cos\psi\Delta_\alpha + j\sin\psi\Delta_\alpha\right), \tag{11}$$

where $i\bar{n}t$ is an interpolation function used in the back-projection operation; $\Delta i = |i - \bar{i}|$, $\Delta j = |j - \bar{j}|$, $\Delta_\alpha = \frac{2\Pi}{\Psi}$, and

$$\sigma^2_\Sigma(i,j) \cong \frac{1}{n_0}\sum_{\psi=0}^{\Psi-1}\sum_n c^2_{n,\theta,ij}\mathrm{e}^{p^h_{n,\theta,ij}}, \tag{12}$$

where n_0 is the initial number of x-ray photons. Thus, it will be possible to find the optimal image μ^* in the sense of estimation of the expected values of the reconstructed image μ. Note, that in Eq. (10) $\tilde{\mu}$ means an image obtained after back-projection operation, $h_{\Delta i,\Delta j}$ are constant coefficients of the convolution, $f(\bullet)$ is a penalty function. We propose the following form of the function $f(\bullet)$:

$$f(e(i,j)) = \lambda \cdot \ln\cosh\left(\frac{e(i,j)}{\lambda}\right). \tag{13}$$

where: λ is a slope coefficient. It is worth emphasizing strongly that the introduction of this function instead of the quadratic amplification of the error in each pixel is not inconsistent with the main idea of the statistical reconstruction approach – that of matching an appropriate divergence function with the probabilistic distribution present in the measured signals. The form of function (13) overlaps with the quadratic form in the wider neighbourhood of their minimums.

Formula (9) can be the basis for the formulation of many solutions to the image reconstruction from projections problem, consistent with ML methodology. Although, there are several methods of searching for the optimal solution, we propose the gradient descent method. In this case, the pixels in reconstructed image will be adjusted in the following way:

$$\mu^{*(t+1)}(i,j) = \mu^{*(t)}(i,j) - c \cdot \sum_{\bar{i}=1}^{\mathrm{I}}\sum_{\bar{j}=1}^{\mathrm{J}}\frac{1}{\sigma^2_\Sigma(\bar{i},\bar{j})} f'\left(e^{(t)}(\bar{i},\bar{j})\right) h_{\Delta i,\Delta j}, \tag{14}$$

where f' is a deriviation of the function (13).

III. Experimental results

In our experiments, we have adapted the well-known Shepp-Logan phantom of the head (see e.g. [11]) for 3D spiral cone-beam projections. Parematers: λ was set to be 2.5, and R_f to be 1200. During the simulations, we fixed $\mathrm{H} = 1025$ measurement points (detectors) on the screen at virtual parallel projections, and $\mathrm{K} = 121$. The number of projections was chosen as $\Theta = 3220$ rotation angles, and the size of the processed image was fixed at $\mathrm{I} \times \mathrm{J} = 1024 \times 1024$ pixels. The coefficients $h_{\Delta i,\Delta j}$ were precomputed before we started the reconstruction process and these coefficients were fixed for the subsequent processing. We started the actual reconstruction procedure and perform the back-projection operation to get a blurred image of the x-ray attenuation distribution in a given cross-section of the investigated object. The image obtained in this way was then subjected to a process of reconstruction (optimization) using an iterative statistically-tailored procedure. It is worth noting that we can choose the starting point of this procedure to be a result of using any standard reconstruction method, for example a reconstruction Feldcamp-type FBP algorithm. Because the set of possible states of matrix μ^* is convex

and the function from relation (3) is convex, the optimization process starting from any point of the convex set μ^* yields a unique solution. The convolutions in iterative procedure were calculated in frequency domain to accelerate reconstruction procedure. The whole iterative process was implemented for GPUs with the NVIDIA CUDA framework and executed on the GeForce GTX 680 graphics card. In this case, the iterative reconstruction process (3000 iterations) has taken 53s.

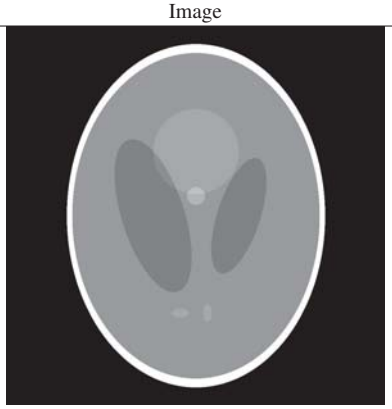
Views of the reconstructed images of the mathematical phantom in the cross-section after 3000 iterations are presented (Table I.c for stochastic signals. For comparison, the original phantom image (Table I.a) and the image reconstructed by a standard Feldkamp-type FBP reconstruction method (Table I.b) are also presented.

IV. Conclusion

We have shown in this paper fully feasible 3D statistical reconstruction algorithm. The use of analytical scheme of signal processing allows us to avoid very serious difficulties associated with the algebraic reconstruction technique, which are particularly noticeable in reconstruction algorithms for spiral scanners. Globally, our reconstruction problem is very easy and compact. There is not used any additional geometrical correction of the projection lines in this approach. Elements of the matrix $h_{\Delta i,\Delta j}$ (see Eq. 11) can be calculated before we start the reconstruction procedure and they are the same for all pixels in the reconstructed image. Simulations have been performed, which prove that our reconstruction method is extremely fast (the whole iterative reconstruction process with 3000 iterations takes 53s) and stable without introducing any additional regularization term, in contrast to the maximum *a posteriori* probability (MAP) estimation. We have achieved this thanks to the introduction of a new form of penalty function in the form of a loss function which prevents the occurrence of possible instabilities in the reconstruction process. The image of the cross-section of the mathematical phantom was reconstructed with high accuracy compared with the standard method (in our experiments the Feldkamp-type algorithm), as measured both in the subjective way and using the objective quality measure.

TABLE I

View of the images (window centre $C = 1.0$, window width $W = 0.2$) when the signals registered by the detectors are deterministic: original image (A); reconstructed image using the standard FBP with Shepp-Logan kernel (B); reconstructed image using the method described in this paper at $t = 3000$ (C).

	Image	MSE
A		—
B		0.0082
C		0.0051

References

[1] A. C. Silva, H. J. Lawder, A. Hara, J. Kujak, W. Pavlicek, *Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm*, American Journal of Roentgenology, vol. 194, pp. 191–199, 2010.

[2] K. Sauer, C. Bouman, *A local update strategy for iterative reconstruction from projections*, IEEE Transactions on Signal Processing, vol. 41, No. 3, pp. 534–548, 1993.

[3] J. -B Thibault, K. D. Sauer, C. A. Bouman, J. Hsieh, *A three-dimensional statistical approach to improved image quality for multislice helical CT*, Medical Physics, vol. 34, No. 11, pp. 4526–4544, 2007.

[4] Y. Zhou, J. -B Thibault, C. A. Bouman, K. D. Sauer, J. Hsieh, *Fast model-based x-ray CT reconstruction using spatially non-homogeneous ICD optimization*, IEEE Transactions on Image Processing, vol. 20, No. 1, pp. 161–175, 2011.

[5] B. DeMan, S. Basu, *Distance-driven projection and backprojection in three dimensions*, Physics in Medicine and Biology, vol. 49, pp. 2463–2475, 2004.

[6] R. Cierniak, *A novel approach to image reconstruction from projections using Hopfield-type neural network*, Lecture Notes in Artificial Intelligence 4029, pp. 890–898, 2006.

[7] R. Cierniak, *A new approach to tomographic image reconstruction using a Hopfield-type neural network*, International Journal Artificial Intelligence in Medicine, vol. 43, No. 2, pp. 113–125, 2008.

[8] R. Cierniak, *A new approach to image reconstruction from projections problem using a recurrent neural network*, International Journal of Applied Mathematics and Computer Science, vol. 183, No. 2, pp. 147–157, 2008.

[9] R. Cierniak, *New neural network algorithm for image reconstruction from fan-beam projections*, Neurocomputing, vol. 72, pp. 3238–3244, 2009.

[10] R. Cierniak, *A three-dimensional neural network based approach to the image reconstruction from projections problem*, Lecture Notes in Artificial Intelligence 6113, pp. 505–514, 2010.

[11] A. C. Kak, M. Slanley, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1988.

Analytical or Monte Carlo system response for pinhole SPECT reconstruction?

Pablo Aguiar, Francisco Pino, Domènec Ros, Javier Pavía and Ziad El Bitar,

***Abstract*—In recent years, many papers on the different methods of obtaining the system response for SPECT reconstruction have been published related. The present work compares the system response obtained by means of analytical and Monte Carlo calculations. To this end, the projections of point sources placed at different distances from a pin-hole collimator obtained by using the analytical and Monte Carlo projectors were compared with experimental data. Our findings show a good agreement between the results obtained with the Monte Carlo method and those using the analytical method including an experimental characterization of the system response. A good agreement was also obtained between the PSF calculated by means of the two projectors and that determined experimentally.**

***Index Terms*—System response matrix, pinhole SPECT reconstruction, Monte Carlo, Raytracing.**

I. Introduction

SMALL animal pinhole SPECT has become an essential tool in preclinical research [1] because of its capability to provide images of peptides, antibodies and hormones. The diffusion of these molecules is slow, thus enabling us to obtain images of processes such as cell division, infection and inflammation. Furthermore, the advantage of the pinhole over parallel-hole collimators is that very high spatial resolution images can be obtained when the object is positioned close to the pinhole. This advantage comes at the expense of a reduced field of view which although it restricts the clinical use of these imaging system, it continues to be suitable for preclinical imaging. The use of pin-hole collimators allows us image quantification in small animals with submillimetric spatial resolution. This high spatial resolution performance should be combined with a degree of sensitivity similar to that found in human studies.

In this regard, statistical reconstruction methods [2], [3], play a vital role in the image quality because they can include accurate descriptions of the acquisition process leading to images of a higher high spatial resolution [4]. The key element of these statistical reconstruction algorithms is the calculation of the system response, which relates the image and the projection space. In recent years many papers have been published related to different methods of obtaining the system response and basically one can say that the modelling can be carried out analytically or by using Monte Carlo simulations.

If the modelling of the system response is carried out analytically, different ray-tracers [5], [6], [7] can be use to compute the geometrical component. Other effects such as intrinsic spatial resolution, finite aperture effect and septal penetration are computed separately [8], requiring a laborious characterization of the system through experimental measurements [9], [10]. Once these effects are modeled, the analytical computation of the system response is fast and it can be obtained *on-the-fly* during the reconstruction process.

As an alternative to the experimental characterization, Monte Carlo simulations can be also used to compute the complete system response (not only the geometrical component) without experimental measurements. The robustness of these techniques in SPECT for parallel [15] and pinhole collimators [16] has been proved.

So far, it still remains unclear which of the two methods for computating the system response is more suitable for pinhole SPECT reconstruction. The aim of the present work is to compare the system response obtained by means of analytical and Monte Carlo calculations.

II. Methods

A. Pinhole SPECT scanner

Our work focuses on a pinhole SPECT system [11] which consists of a rotating flat detector with a monolithic scintillator crystal (Cesium Iodine, CsI) of $50 \times 50 \times 4mm^3$ coupled to H8500 PS-PMT and $1\ mm$ diameter tungsten pinhole [12]. The gamma camera is attached to a variable radius system in order to optimize sensitivity and resolution by adjusting the radius of rotation to the size of the object.

B. Analytical system response and experimental PSF

The computation of the different elements of the system response matrix can be a very tedious task, because of the large size of the matrix. To facilitate the computation, the complete matrix was factorized into several submatrices [8],

P. Aguiar is with Molecular Imaging Group, Nuclear Medicine Department, Complexo Hospitalario Universitario de Santiago de Compostela and IDIS/IDICHUS, Spain,

F. Pino are with Unitat Biofísica, Universitat de Barcelona and Servei de Física Mèdica i Protecció Radiològica, Institut Català d'Oncologia, L'Hospitalet de Llobregat, Spain.

D. Ros are with Unitat Biofísica, Universitat de Barcelona and CIBER Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain.

J. Javier Pavía is with Hospital Clínic de Barcelona and CIBER Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain

Z. El Bitar is with Imabio, Institut Pluridisciplinaire Hubert Curien (IPHC), CNRS, UMR 7178-CNRS/IN2P3 and UDS, Strasbourg, France

each one related to a relevant aspect in the image formation process.

1) Geometrical component: the geometrical component of the system response was obtained by using a fast Siddon ray-tracer [5], so that for each bin-voxel element, the response was computed as the length of intersection between the line related to the bin and the voxel.

2) Intrinsic spatial resolution: in addition to the geometrical component, PSF effects related to the intrinsic spatial resolution were also included. A characterization of the intrinsic spatial resolution was carried out through experimental measurements from point sources using a pencil beam and a set of detector intrinsic PSFs was obtained for different incidence angles.

3) Finite aperture and septal penetration: analytical models of finite aperture and septal penetration were also included. To this end, we considered an equivalent aperture diameter yielding a geometric resolution equal as the total resolution of the real pin-hole [13]. The scatter in the collimator was not considered.

C. Monte Carlo system response

Monte Carlo simulations were used to obtain an accurate description of all the phenomena involved in the SPECT imaging process which relate the activity distribution and the photons collected by the detector.

1) GEANT4 simulations: the complete system response was obtained by using GEANT4 simulations [14] of a cylinder with uniform activity concentration [15], [16]. Simulation took into account most of the physical phenomena involved in the radiation-matter interaction, including the gamma emission, pinhole collimation and gamma detection. Thus, photoelectric absorption, Compton scattering and Rayleigh scattering were activated in the simulations. Tungsten characteristics x-rays were not included. To tackle the hindrance of large computing time, parallel simulations were performed on a computing grid composed on 1000 computing elements.

2) Variance reduction: variance reduction methods based on forced detection were needed to obtain noise-free simulations within reasonable time. Thus, accelerated Monte Carlo simulations [17] were carried out by forcing the direction of photons towards a circular area centered at the pin-hole collimator. The detected photons were weighted to compensate the fact that emission was constrained within a limited solid angle.

D. Mechanical misalignments

The mechanical misalignments and system response modeling are closely related topics in pinhole SPECT imaging, and a reconstruction method implemented without the capability of correcting for geometric misalignments may be of little interest in practical molecular imaging studies. That is why the reconstruction of pinhole SPECT data requires a correct description of the acquisition geometry and this information has to be included in both the analytical calculations and Monte Carlo simulations. These mechanical misalignments were completely described taking into account the radius (distance between the center of the hole and the center of rotation), the focal length (distance between the center of the hole and the detector), the mechanical offset (distance between the rotation axis and the pin-hole projection ray orthogonal to the detector), the tilt angle (angle between the detector and the rotation axis), the twist angle (describing a rotation of the detector around an axis parallel to the central ray) and the electrical shifts (describing a collective translation of the projection image caused by drift of the detector hardware) [18].

It remains unclear if other uncertainties such as pinhole opening angle ($90\ deg$), the pinhole orientation or the pinhole height ($2\ cm$) can affect the image quality. Future investigations in this regard would be of great interest.

E. Calibration

The calibration parameters were obtained from a sixty projections acquisition of a three-point 99mTc phantom [18] and minimizing a penalty function defined as the sum of quadratic differences of the coordinates of the projected centers of the point sources and the experimental acquired centers.

F. Experimental and simulated images of point sources

A comparison of the system response models in terms of PSF was carried out. To this end, point sources were projected by using analytical (including experimental characterization) and Monte Carlo system response. PSF parameters were obtained from profiles of images of a point source placed at different distances and angles of the pinhole collimator. The point sources were placed at distances greater ($26.4\ mm$, $27.2\ mm$ and $29.1\ mm$) and smaller ($16.6\ mm$, $17.9\ mm$ and $18.1\ mm$) than the radius of rotation, and for three different incidence angles (4, 13 and 20 deg). Experimental PSFs were also obtained for comparison.

III. Results and discussion

A. Calibration parameters

The geometrical parameters obtained from the calibration acquisition by using the three-point 99mTc phantom were included in both analytical and Monte Carlo system response modeling. The estimated radius was $22.5\ mm$ and focal distance was $33.1\ mm$, thus showing a magnification factor of 1.47. With respect to the parameters related to the detector, the mechanical offset was $3.1\ mm$, tilt angle was $0.75°$, twist angle was $0.69°$ and the electrical shifts were $-4.5\ mm$ and $2.3\ mm$, for transverse and axial directions respectively. These results suggest that the calibration of the mechanical offset and the electrical shifts can be essential for this acquisition.

Parameter	Value
Radius of rotation	22.5 mm
Focal length	33.1 mm
Mechanical offset	3.1 mm
Tilt angle	0.75 deg
Twist angle	0.69 deg
Transverse shift	-4.5 mm
Axial shift	2.3 mm

TABLE I
CALIBRATION PARAMETERS

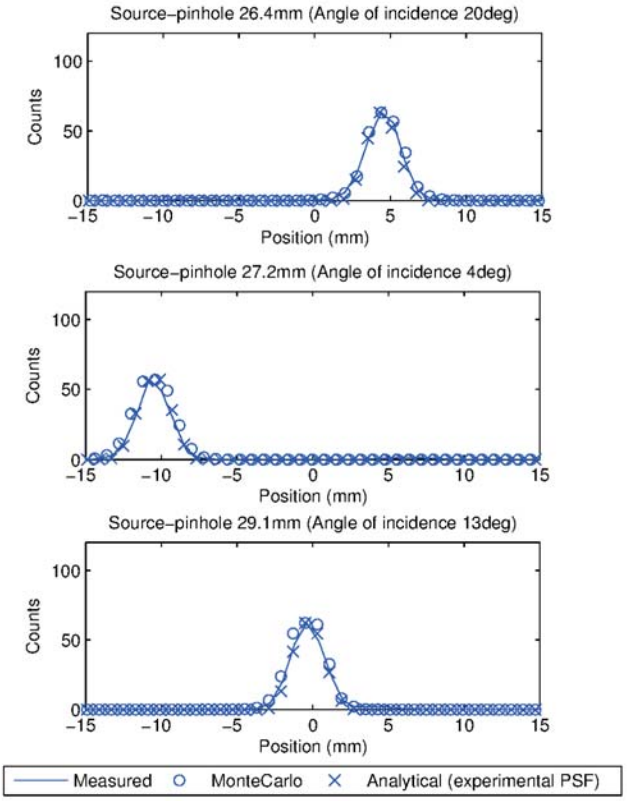


Fig. 1. System response for point sources placed at distances greater than the radius of rotation ($> 22mm$) for different angles of incidence ($4°$, $13°$ and $20°$). The transverse profiles were obtained through projected points by using Monte Carlo and analytical system response modelling.

B. Comparison between system response models

Figure 1 and Figure 2 show the transverse profiles obtained through projected points by using Monte Carlo and analytical system response modeling, and also the transverse profiles from experimental measurements for comparison. It can be observed that the projection for small angles of incidence (for example $4°$) corresponds to a point at the edge of the detector, not at the center of the detector, as expected. On the other hand, point sources are detected at the center of the detector for angles of incidence of $13°$. These findings can be explained by the mechanical misalignments, particularly the mechanical offset and electrical shifts (tilt and twist angles were quite small $< 1°$). It is noteworthy that an excellent agreement exists in the position of each projected point sources (measured, Monte Carlo and analytical), thus

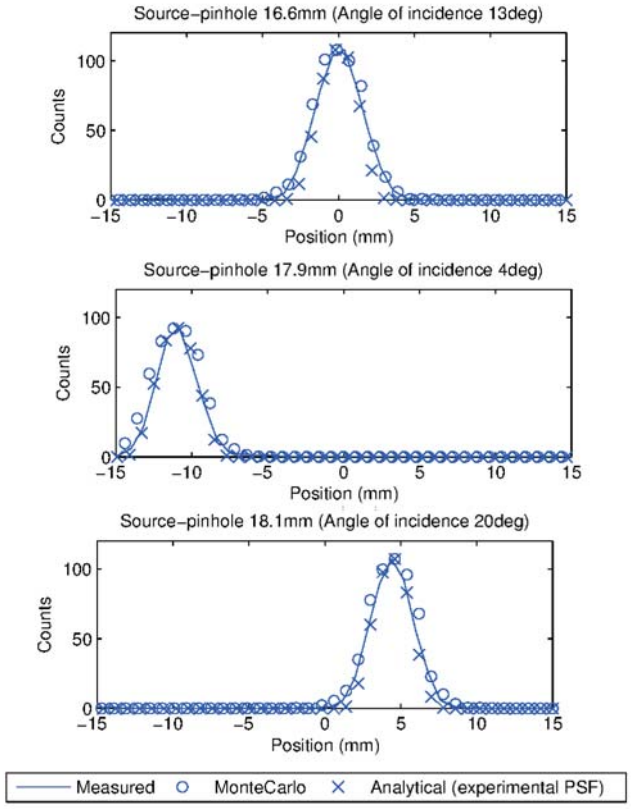


Fig. 2. System response for point sources placed at distances smaller than the radius of rotation (< 22 mm) for different angles of incidence ($4°$, $13°$ and $20°$). The transverse profiles were obtained through projected points by using Monte Carlo and analytical system response modelling.

demonstrating the correct implementation of the different calibration parameters.

Figure 1 shows the system response for point sources (PSFs) placed at distances greater than the radius of rotation (> 22 mm) for different angles of incidence ($4°$, $13°$ and $20°$). The plots of Figure 1 indicate that, in general terms, the system response modeling by using analytical and Monte Carlo is in good agreement with measured data, thereby demonstrating the accuracy of both approaches.

Figure 2 shows the system response for point sources (PSFs) placed at distances smaller than the radius of rotation (< 22 mm) for different angles of incidence ($4°$, $13°$ and $20°$). In general terms, a good agreement between simulated and experimental data is also observed.

A detailed analysis of the PSFs shows some relevant issues that should be mentioned. Thus, experimental and simulated PSFs obtained from point sources placed at distances greater than the radius of rotation (27.2 mm, 29.1 mm and 26.4 mm) show a better agreement than those obtained from point sources placed at distances smaller than the radius of rotation (17.9 mm, 16.6 mm and 18.1 mm). In this case, the PSFs obtained from the analytical modeling were slightly narrower

than those corresponding to experimental measurements, whereas PSFs obtained from the Monte Carlo modeling were slightly wider than those obtained experimentally. The latter is particularly relevant for an angle of incidence of 4°, which corresponds to a point at the edge of the detector.

Due to the fact that only slight differences were found between the PSFs obtained using the analytical and Monte Carlo system response, in the election of one or other approaches should take into account the experience in these fields. The analytical system response requires an experimental characterization to achieve a performance similar to that obtained by using Monte Carlo system response, but the latter requires large calculation times and an extensive background in the use of Monte Carlo simulation tools.

IV. Conclusion

A comparative study between two system response models has been carried out for pinhole SPECT reconstruction in terms of PSFs.

A good agreement was found between the results obtained with the Monte Carlo method and those using the analytical method including an experimental characterization of the system response. A good agreement was also obtained between the PSFs calculated by means of the two projectors and those determined experimentally, although slight differences were found for point sources close to the pinhole. Further studies are being conducted to assess the effect of these slight differences on the quality of the reconstructed images.

Acknowledgments

This work was supported in part by Fondo de Investigaciones Sanitarias del Instituto de Salud Carlos III Project Nos. PS09/01206 and PI11/01806 and Xunta de Galicia 10CSA918001PR. Pablo Aguiar was awarded a Sara Borrell fellowship by Fondo de Investigaciones Sanitarias del Instituto de Salud Carlos III. The authors also acknowledge the support of IPHC for providing computing resources on the Regional Grid Infrastructure (http://www.grand-est.fr).

References

[1] Meikle SR, Kench P, Kassiou M, Banati RB.“Small animal SPECT and its place in the matrix of molecular imaging technologies”. Phys Med Biol **50** (22): R45R61 (2005)

[2] Shepp L and Vardi Y, “Maximum-likelihood reconstruction for emission tomography”, IEEE Trans Med Imaging **1**, 113-121 (1982).

[3] Hudson H and Larkin R, “Accelerated image reconstruction using ordered subsets of projection data”, IEEE Trans Med Imaging **13** (4), 601-609 (1994).

[4] Frese T, Rouze N, Bouman C, Sauer K and Hutchins G, “Quantitative comparison of FBP, EM and Bayesian reconstruction algorithms for the IndyPET scanner”, IEEE Trans Med Imaging **22** (2), 258-276 (2003).

[5] Siddon R, “Fast calculation of the exact radiological path length for a three-dimensional CT array”, IEEE Trans Med Imaging **12**, 252257.

[6] Wu X, “An efficient anti-aliasing technique”, ACM Computer Graphics Siggraph Conf Proc, **4** (25), 143-152 (1991).

[7] Schretter C, “A fast tube of response ray-tracer”, Med Phys **33** (12), 4744-4748 (2006).

[8] Qi J, Leahy R, Chatziioannou A, Cherry S and Farquhar F, “High-resolution 3D Bayesian image reconstruction using the microPET small animal scanner”, Phys Med Biol **43** (7), 1001-1013 (1998).

[9] van der Have F, Vastenhouw B, Rentmeester M and Beekman FJ. “System calibration and statistical image reconstruction for ultra-high resolution stationary pinhole SPECT”. IEEE Trans Med Imaging **27** (7):960971 (2008)

[10] Feng B, Chen M, Bai B, Smith AM, Austin DW, Mintzer RA, Osborne D and Gregor J. “Modeling of the point spread function by numerical calculations in single-pinhole and multipinhole spect reconstruction”. IEEE Trans Nucl Sci **57** (1):173180 (2010)

[11] Pino F, Roe N, Orero A, Falcon C, Rojas S, Benlloch JM, Ros D and Pavia J. “Development of a variable-radius pinhole SPECT system with a portable gamma camera”. Rev Esp Med Nucl **30** (5):286-291 (2011)

[12] Fernandez MM, Benlloch JM, Cerda J, Escat B, Gimenez EN, Gimenez N, Lerche CW, Martinez J, Pavon N, Sanchez F and Sebastia A. “A flat-panel-based mini gamma camera for lymph nodes studies”. Nucl Instr and Meth Phys A **527**:92-96 (2004)

[13] Accorsi R, Metzler SD. “Analytic determination of the resolution-equivalent effective diameter of a pinhole collimator”. IEEE Trans Med Imaging, 23(6), 750-63.

[14] Agostinelli et al. “Geant4 - a simulation toolkit”. Nucl Instr and Meth Phys A **506**:250-303 (2003)

[15] Lazaro D, El Bitar Z, Breton V, Hill D and Buvat I. “Fully 3D Monte Carlo reconstruction in SPECT: a feasibility study”. Phys Med Biol, **50** (16):3739-3754 (2005)

[16] El Bitar Z, Lazaro D, Coello C, Breton V, Hill D and Buvat I. “Fully 3D Monte Carlo image reconstruction in SPECT using functional regions”. Nucl Instr and Meth Phys A **569** (2):399-403 (2006)

[17] El Bitar Z, Breton V, Hill D and Buvat I. “Acceleration of fully 3D Monte-Carlo based system matrix computation for image reconstruction in small animal SPECT”. IEEE Trans Nucl Sci **58** (1):121-132 (2011)

[18] Beque D, Nuyts J, Bormans G, Suetens P and Dupont P.“Characterization of pinhole SPECT acquisition geometry”. IEEE Trans Med Im, 22(5), 599612 (2003)

Strategies for GPU-based cone-beam CT reconstruction for very large data volumes

Xuan Liu, Stephan Boons, Alexander Sasov

***Abstract*—The rapid developments in technologies of microCT and nanoCT bring forth large amounts of data. Reconstruction often forms a bottleneck in data processing. Hardware acceleration has become a hot topic, especially the usage of Graphics Processing Unit (GPU). Using the C-like language CUDA provided by NVIDIA, we have developed a GPU-based reconstruction package, which can run with multiple GPU's on one or more PC's. This paper describes and discusses the strategies we use to deal with reconstruction of very large data volumes on GPU's with limited on-board memory. The reconstruction algorithm is the commonly used FDK algorithm for circular orbit cone-beam CT. We attempt to achieve the best overall performance. Therefore, not only the back projection kernel on GPU's needs to be optimized, but also the task division and the data flow, between hard drive and CPU memory and between CPU memory and GPU memory, need great attention. Evaluation of this software package on up-to-date CUDA-enabled GPU's is performed for datasets up to, but not limited to, 675 Giga voxels (15Kx15Kx3K).**

***Index Terms*—Cone-beam CT, image reconstruction, multi-GPU, distributed, CUDA, FDK algorithm**

I. Introduction

The aim of GPU-based reconstruction project was to provide our microCT and nanoCT users with an affordable and convenient solution to reconstruction acceleration. The algorithm in use is the commonly used Feldkamp[1], which is still the golden standard for microCT reconstruction. The good performance of CUDA-based GPU reconstruction demonstrated by many authors in recent years [2,3,4], the widespread usage of NVIDIA GPU's and their fast and steady development, and the simplicity and maturity of CUDA framework and tooling have encouraged us to choose a CUDA-based GPU implementation.

One particular challenge for this project is the size of the data volumes. The CUDA-enabled NVIDIA GPU's have on-board memory ranging from 256MB to 6GB[5], the majority has 1~3GB memory. The data volume we are dealing with goes from 512^3 to 3000 x 15000^2 voxels, resulting in 0.5GB to 2700GB in floating point format. Most authors [2, 3, 4] divide an image volume into sufficiently smaller sub-volumes so that these can be kept on GPU memory during one back-projection sub-task. This will become highly inefficient when the sub-volumes are reduced to a few slices in case of extreme large datasets relative to the available GPU memory. Such inefficiency is caused by overhead intrinsic to cone-beam geometry, as explained in section II.A.

A "very large data volume" can be understood as: neither the complete input necessary for reconstructing a sub-volume, nor the full reconstruction volume can be put into CPU memory; let alone the smaller and scarcer on-board GPU memory. This implies that, for a CUDA back-projection sub-task, neither the entire input, nor the entire output can be kept on board. We solve this problem by using a sufficiently large slice-buffer on CPU, which is updated slice by slice while GPU back-projects views subset by subset. This scheme works efficiently if the buffer sizes on CPU and GPU are carefully chosen to guarantee efficient data flow and the data transfer time between CPU and GPU can be hidden sufficiently.

This paper presents strategies of data management and task divisions in reconstructing large data volumes in a distributed multi-GPU system.

All authors, Xuan Liu (e-mail: xuan.liu@bruker-microct.com), Stephan Boons (e-mail: stephan.boons@bruker-microct.com), and Alexander Sasov(e-mail: alexander.sasov@bruker-microct.com) are with Bruker-microCT, Kartuizersweg 3B, 2550 Kontich, Belgium.

II. GPU-based reconstruction

A. The FDK algorithm

The FDK algorithm consists of three independent steps: pre-processing, filtering and back-projection. The pre-processing step includes ring artifact correction, beam-hardening correction, smoothing and applying logarithmic conversion.

In our GPU implementation, the pre-processing is still done on CPU, whereas the filtering and back-projection have been ported to GPU.

This algorithm is well suited for parallel programming. We divide a full image volume into smaller chunks; each contains a section of contiguous slices. These sub-volumes can be reconstructed completely independently. There is, however, due to the cone-beam geometry, an overlapping in pre-processing and filtering of the projection lines between adjacent chunks, as shown in fig. 1. The overlapping becomes larger when the cone angle increases and is the largest for sub-volumes at bottom/top. The overhead becomes significant when image volume is very large comparing to available memory, and the number of slices in one chunk becomes smaller. In CPU implementation, we use disk as an extended buffer to keep filtered projections, so that the overlapped projection sections do not need to be redone. In GPU implementation, this is no longer an option, as the disk

access becomes much slower than reconstruction. Fortunately the filtering time on GPU with CudaFFT is almost negligible compared to back-projection. Therefore, we have chosen to repeat the pre-processing and filtering of overlapping sections whenever necessary, while maximizing image buffer on CPU. To avoid network and disk latency caused by reading, a relatively large projection buffer in CPU memory is used.

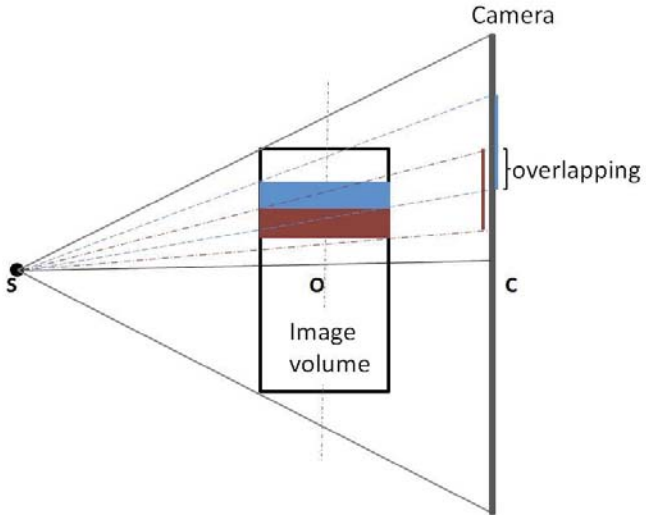


Fig. 1. Scheme of task division. The full image volume is divided in vertical direction into sub-volumes, which can be reconstructed independently. Due to the cone-beam geometry, the areas on camera required for reconstructing the sub-volumes do overlap.

B. Distributed multi-GPU architecture

Mainstream PC's can host one or two GPU's, while dedicated hardware allows up to 16 GPU's. When dedicated hardware is not available, a distributed version, similar to the CPU counterpart [6] of our software, is a logical way to use more GPU's in one setup, as shown in fig. 2. For best performance, the PC's, referred as nodes, should be within a local network connected by Gigabit switch or InfiniBand. A client-server system with standard socket communication is used: the client program has a user interface for configuring the nodes, setting the reconstruction parameters and managing the communication with servers running on nodes, whereas a server program on each node receives sub-tasks and performs reconstruction using GPU's available to it. The coordination is done by the client and the nodes work independently from each other.

The nodes do not need to be identical, as the sizes of the chunks are calculated according to the capacities of the nodes. On each node, the task will be further divided into sub-tasks among the GPU's, as described in section D. The GPU's on each node should be identical for optimal performance.

C. Optimization on GPU

We have implemented a reconstruction library using CUDA. Both forward-projector and back-projectors are included. For this paper, only the back-projector is used. The voxel-driven method is used for back-projection.

Many authors have worked on optimization of back-projection using CUDA [2, 3, 4, 7]. The conventional wisdom can be summarized as the following: 1) Access to global memory is very expensive (>400 cycles), so this should be kept to a minimum. If it cannot be avoided, improving access locality can help. 2) If possible, use the read-only texture memory: it has not only the much more efficient memory read due to caching, but also hardware-assisted fast interpolation.

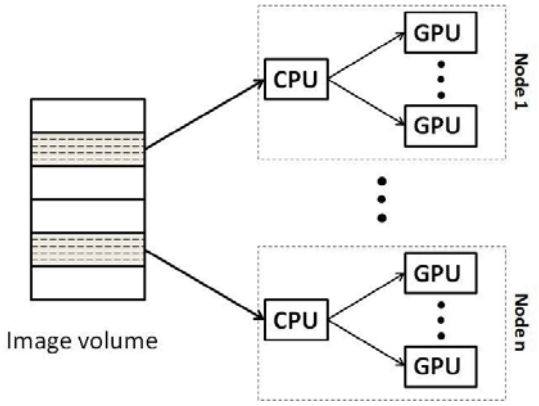


Fig. 2. Distributed multi-GPU architecture. One reconstruction task is distributed among several nodes (PC's); each of them is equipped with 1 or more GPU's.

Based on these valuable observations, we have chosen to keep projections in texture memory and the image volume in the global linear memory. By using M (>=8) projections within one back-projection kernel, the required number of global memory access is reduced by a factor of M, as the voxel value can be kept in register during back-projection. This speeds up the back-projection kernel up to 3 times, as shown in fig. 3.

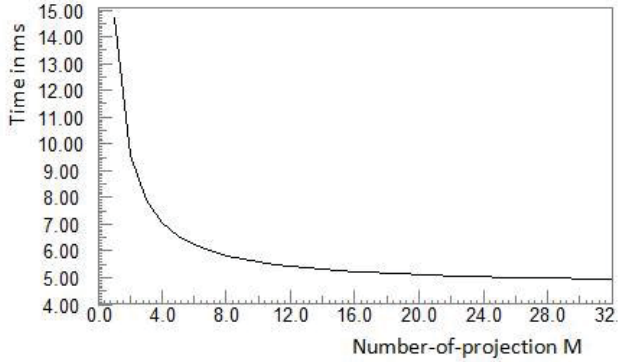


Fig. 3. Back-projection speed in function of number of projections M. This test was done on a Tesla C2075 card. M projections of 8Kx1K were back-projected to 50 8Kx8K slices. The times shown are normalized to milliseconds per projection per slice. Data transfer time between CPU and GPU was not included.

Due to limited GPU memory, the projections and image slices have to be partitioned. To find the most optimal partitioning, we have investigated two different partitioning schemes: multiple-projections and single-slice (MPSS), few-projections and multiple-slices (FPMS). Fig. 4 and fig. 5 show the pseudo-code for each scheme.

In the MPSS scheme, only one slice (two 1-slice buffers are allocated if asynchronous transfer is used) is kept on GPU. The number of projections M is then maximized according to the available GPU memory. The corresponding projection sections are concatenated in one 2D texture. Layered 2D texture would be a better choice, but was not used for the time

being, due to dimension limitations imposed by CUDA. One limitation of a 2D texture is the maximum height (32768 lines), which has to be taken into account when determining M. During back-projection of each subset, each of the S slices needs to be transferred back to CPU. We avoid uploading of slices by doing the summation in CPU in a separate thread while GPU is busy with the next slice. In principle, the data transfer time can be hidden on most devices by overlapping data transfer and kernel execution. This is efficient as long as the transfer time does not exceed the calculation time. This condition is well satisfied if M is sufficiently large. Taking a 8Kx8K dataset running on Tesla C2075 as an example: when M=32, the speed is 0.312ms per projection per slice, thus the total back-projection time of a subset is 32*0.312=9.984ms; data transfer rate between CPU and GPU is slightly more than 5GB/s, giving a transfer time of 2.5ms. As M increases, the proportion of data transfer time in the total time drops proportionally. This is another reason why we want to maximize M; even though the kernel speed doesn't increase significantly beyond M= 8. In practice, M is usually larger than 100 so that the transfer time is almost negligible.

```
// Scheme Multi-projections and single-slice (MPSS)
For each projection-subset (N subsets)
    Upload the M projections in the subset to GPU
    Apply filtering
    For each slice (S slices)
        Zero slice-buffer
        For each voxel (X x Y voxels)
            Back-project from the M projections in the subset
        End for each voxel
        Transfer slice back to CPU for summation
    End for each slice
End for each projection-subset
```

Fig. 4. Pseudo-code for back-projection scheme MPSS.

```
// Scheme Few-projections and Multi-slices (FPMS)
Zero S-slice buffer
For each projection-subset (n subsets)
    Upload m projections in the subset to GPU
    Apply filtering
    For each slice (S slices)
        For each voxel (X x Y voxels)
            Back-project m projections in the subset
        End for each voxel
    End for each slice
End for each projection-subset
Download all S slices back to CPU
```

Fig. 5. Pseudo-code for back-projection scheme FPMS.

In the FPMS scheme, the number of slices on the GPU is maximized according to available memory, while guaranteeing a buffer of minimum 8 projections. The major advantage of this scheme is that the sub-volume stays on the GPU during the entire back-projection.

In our implementation, we choose one of the two schemes automatically according to data volume and the available CPU/GPU memory. FPMS scheme is rarely in action due to its excessive demand on GPU memory, whereas MPSS scheme is always applicable as long as the GPU memory is large enough to hold one slice.

D. Data flow and task division between CPU and GPU's

To achieve good overall performance, it is crucial to avoid bottlenecks during the full cycle of reconstruction: from reading projections to writing results back to disk. Not only the uploading to and downloading from a GPU need to be scheduled carefully, but also the reading and writing require more attention, because the disk bandwidth is usually much lower than that for GPU.

Similar to the task division on the level of nodes, the reconstruction volume on a single PC is further divided in chunks. The maximum number of slices contained in one chunk is determined by available CPU memory. Fig. 6 shows the scheme for reconstructing one such chunk. To hide disk latency, a reading thread manages the reading to the buffer containing P projections; in the meanwhile, another thread pre-processes the available projections. We set P to be larger than M to hide potential reading/pre-processing delays. Each GPU is managed by one CPU thread, which coordinates data transfer to/from GPU, and synchronizes filtering and back-projection processes on GPU. To hide the disk latency during writing, a writing thread monitors the status of each slice and the writing of a slice is started as soon as it is ready. The writing continues while the reading, pre-processing and filtering are going on for the next chunk.

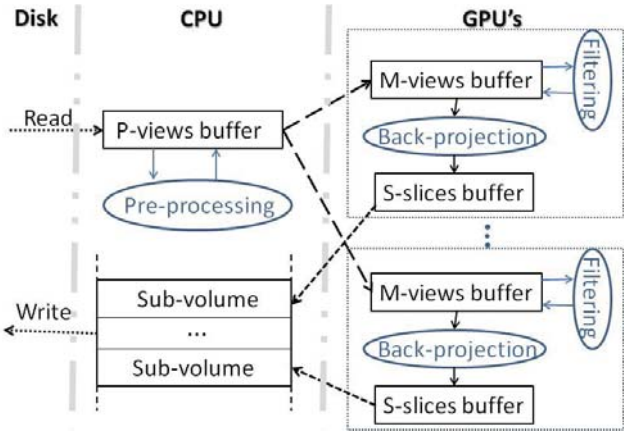


Fig. 6. Reconstruction scheme with a single CPU equipped with one or more GPU's. CPU maintains a buffer with P projections which are read, pre-processed and sent to each GPU, and an image volume which will be reconstructed by GPU's and written to destination disk. Each GPU is controlled by one CPU thread for filtering and back-projection.

III. Evaluation on various hardware

We have evaluated the GPU-based reconstruction on various GPU's, including the most recent GeForce gaming cards, the Quadro cards and the Tesla cards. Our CPU reconstruction software [6] was used as reference. No image degradation was observed for the GPU implementation (results not shown).

To compare speed, the multi-threaded CPU reconstructions were done on our current standard recommended computer, not on the most powerful personal computer. A set of typical datasets with different dimensions are used, as listed in table 1. The largest 15K dataset is not included in the comparison,

as it would have taken very long to complete all tests. Only the CXT8000 with 8 Tesla C2075's could finish the reconstruction in reasonable time: 18.5 hours to reconstruct a volume of 2610 x 15K x 15K from 8100 projections of 15K x 3K.

dataset	1K	2K	4K	8K
#views	499	996	1990	2157
Width	1000	2000	4000	7744
Height	670	1340	2680	2680
#images	615	1229	2225	2495
Size	1024^2	2000^2	4000^2	7744^2

Table1. Dimensions of the test datasets.

A. Single GPU's

Table 2 shows the reconstruction times of the test datasets on individual cards. As the graphics cards come with different sizes and variable power requirement, it is not possible to test them all on a common PC. Attention was paid to the performance of hosting PC's so that the GPU performance was not noticeably degraded by I/O and CPU processing. The speedups range from 4 to 15 times.

	1K	2K	4K	8K
C2075	31s	9m39s	3h3m	8h8m
GTX690[a)]	17s	2m58s	44m23s	3h50m
GTX590[a)]	21s	4m7s	1h2m	4h35m
GTX580	24s	7m0s	1h31m	6h9m
GTX570	29s	9m53s	1h34m	8h0m
FX3800	54s	16m7s	3h35m	16h13m
Quadro4000	58s	17m3s	4h7m	14h57s
CPU[b)]	*3m58s*	*58m*	*15h13m*	*59h12m*

Table 2. Reconstruction times on various GPU's. a) These devices contain 2 units per card; b) For CPU version, a DELL R5500 PC, with 1 X5675 processor (3.07GHz, 6 cores) and 24GB RAM, was used.

B. Multi- GPU's

We have a multi-GPU setup on a CXT8000 with 8 Tesla C2075's. As shown in table 3, the speed increases almost linearly with the number of cards, except for the smallest dataset, where the data transfer becomes more influential for overall performance.

	1K	2K	4K	8K
1xC2075	31s	9m39s	3h3m	8h8m
8xC2075	14s	1m34s	19m0s	1h8m

Table 3. Linearity test results for a multi-GPU setup.

C. Distributed multi-GPU's

To test this configuration, we built a 4-node cluster. The nodes were connected with a GigaBit switch. Each node had a GTX590 card (2 GPU's per card), the fastest single device at the moment of testing. As shown in table 4, we obtain speed advantage up to 3 nodes. The performance degradation starting from 4 nodes is due to the slow network speed, mainly during writing. Alternative mechanism should be searched to further hide the network latency. An InfiniBand instead of GigaBit switch could be a possible hardware solution.

	1K	2K	4K	8K
1 node	21s	4m7s	1h2m	4h35m
2 nodes	16s	2m32s	34m26s	2h26m
3 nodes	16s	1m33s	24m	1h50m
4 nodes	12s	1m24s	24m	1h22m

Table 4. Linearity test results for a distributed multi-GPU setup.

IV. Conclusions

We have developed a distributed multi-GPU FDK reconstruction software package for microCT. We show that FDK algorithm can be implemented efficiently for very large data format for GPU devices with modest on-board memory. For the time being we require the amount of memory on device should be large enough to hold one slice. The limitation can in principle further reduced as it is no problem to divide a slice into sub-slices in voxel-driven back-projection, and it will remain efficient as long as the data transfer between CPU and GPU can be sufficiently hidden.

To obtain good overall performance for large datasets, an optimal data flow has been designed to reduce latencies at different levels. At GPU level, this is done by choosing carefully the reconstruction scheme so that the uploading and downloading time can be either efficiently hidden or reduced to a negligible level. At CPU level, this is done by careful partitioning and synchronization among reading, writing, pre-processing and reconstruction threads, so that reading, pre-processing and writing time can be largely hidden.

From performance point of view, a single PC with multiple GPU's is much more efficient than a distributed system, as reading/writing via network quickly forms a bottleneck for the whole data flow.

Acknowledgment

We thank colleagues at Bruker MicroCT and distributors of Bruker MicroCT for their tests and suggestions. We are also grateful to our users for their great enthusiasm in this project and for their valuable feedback.

References

[1] L.A. Feldkamp, L.C. Davis, and J.W. Kress, "Practical conebeam algorithm," J. Opt. Soc. Am. A, vol. 1, pp. 612– 619, 1984.

[2] S. Sawall, etc., "Performance Comparison of OpenCL and CUDA by Benchmarking an Optimized Perspective Backprojection", Proc. Fully 3D meeting, Potsdam, 2011, pp. 15-18.

[3] E. Papenhausen,etc., GPU{-Accelerated Back-Projection Revisited: Squeezing Performance by Careful Tuning", Proc. Fully 3D meeting, Potsdam, 2011, pp. 19-22

[4] Y. Okitsu, F. Ino and K. Hagihara. "High-Performance Cone Beam Reconstruction Using CUDA Compatible GPUs," Parallel Computing, 36(2-3):129-141, 2010.

[5] http://www.nvidia.co.uk/object/cuda_gpus_uk.html.

[6] X. Liu, and A. Sasov, "Cluster reconstruction strategies for microCT/Nano-CT scanners," Proc. *Fully 3D meeting*, Salt Lake city, 2005.

[7] S.Q. Zheng, etc., "A Distributed Multi-GPU System for High Speed Electron Microscopic Tomographic Reconstruction", Ultramicroscopy. 2011 July ; 111(8): 1137–1143.

Multi-Material Beam Hardening Correction(MMBHC) in Computed Tomography

Qiao Yang, Nicole Maass, Mengqiu Tian, Matthias Elter, Ingo Schasiepen, Andreas Maier, and Joachim Hornegger

***Abstract*—In computed tomography (CT), the nonlinear attenuation characteristics of polychromatic X-rays cause beam hardening artifacts in the reconstructed images. State-of-the-art methods to correct the beam hardening effect are mostly single material precorrections (e.g. water-precorrection), which are far less efficient when more than one material is present in the field of measurement. The use of those techniques is limited by specific restrictions to the objects, computational loads, and inaccurate segmentations. In this paper, we present a practical multi-material beam hardening correction(MMBHC) approach that employs material decomposition technique maintaining CT values from dual-energy CT. This separates single energy CT images into spatial density images and images containing material information. The segmentation maintains the original X-ray attenuation coefficients, such that the original CT attenuation image can be exactly recovered. Therefore, segmentation errors, which result in invalid material properties to the voxel, only have minor effects on the beam hardening correction and do not cause an atypical image impression or introduce additional artifacts. A multi-material beam hardening correction procedure is formulated to iteratively correct the artifacts but shows satisfactory image quality after the first iteration. Based on experiments with simulated CT data, it is shown that the proposed method can efficiently reduce beam hardening artifacts. In addition to the performance benefits, our approach can be flexibly applied to imaging geometries and achieve efficient, fully 3D reconstructions.**

***Index Terms*—CT reconstruction, beam hardening, artifact reduction, segmentation, spatial resolution, image quality**

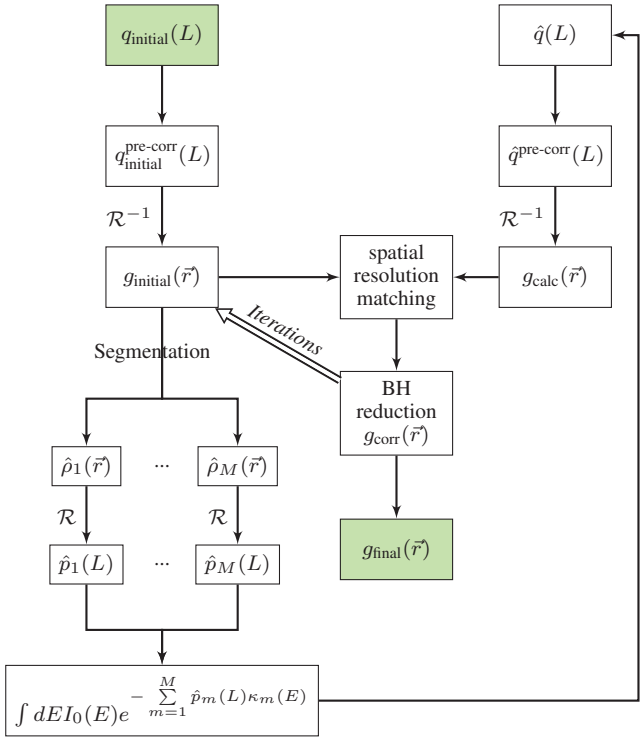


Fig. 1: Flowchart of our proposed algorithm

I. Introduction

In computed tomography (CT), standard reconstruction techniques are generally based on the assumption that the X-ray beams are monochromatic and the measured projection images contain line integrals through the objects. However, in practice, the X-rays are polychromatic and lower energy photons are more easily absorbed than the higher energy photons, resulting in nonlinear characteristics of the X-rays, the so-called beam hardening phenomenon. The polychromatic characteristic of X-rays leads to the attenuation of a homogeneous object not being proportional to the thickness of the object along the ray. Consequently, severe artifacts such as cupping and streak artifacts appear in the reconstruction, which compromise the reconstruction quality and diagnostic accuracy. Therefore, an effective beam hardening correction approach is important in both medical and industrial CT applications to improve image quality.

To mitigate beam hardening, common methods are hardware-based approaches to reduce the polychromaticity of the incident X-rays. State-of-the-art beam hardening correction (BHC) algorithms used in clinical CT are mainly based on single material (e.g. water) calibration and can efficiently correct objects consisting of materials that are spectrally alike [1]. In dual energy CT, the aim is oftentimes to reconstruct images of material densities and as a side-effect, the beam hardening artifacts can be exactly corrected [2]–[5].

In this work, we focus on the single energy CT of objects that consist of multiple materials (with different spectral properties). Prior research on this topic requires a segmentation of the attenuation image from a single energy CT scan into different materials [6]–[14]. A major limitation of these techniques is the computational complexity. Moreover, require previous knowledge, the objects' characteristics such as number of materials, material's inhomogeneity and the shape of different materials also render those techniques unsuitable for practical usage. The high computational load of these methods is often

Q. Yang, Dr. Andreas Maier and Prof. Dr. J. Hornegger are with the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Martenstrasse 3, Erlangen, Germany and the Erlangen Graduate School Heterogeneous Imaging Systems. Dr. Andreas Maier and Prof. Dr. J. Hornegger are with the Erlangen Graduate School in Advanced Optical Technologies (SAOT). Q. Yang, Dr. N. Maass, M. Tian, Dr. M. Elter and I. Schasiepen are with the Siemens AG, Healthcare Sector, Erlangen, Germany. This work was partly supported by the Research Training Group 1773 "Heterogeneous Image Systems", funded by the German Research Foundation (DFG).

Corresponding author: qiao.yang@cs.fau.de.

caused by missegmentation of materials in early iterations, which slows down convergence and may result in atypical image impression and additional artifacts.

We present an application of the dual-energy reconstruction technique presented in reference [5] with a sophisticated segmentation method which makes it applicable to single energy CT. Both methods calculate polychromatic raw data from material density images of base materials. While the calculation of base material density images is the aim of dual energy CT, we use maintained CT attenuation values for segmentation in single energy CT to generate material density images. Thereby, the CT attenuation value conservation leads to a situation where the consequences of missegmentation in the beam hardening correction are minor. We do not make any assumptions on the scanned object, but we assume to know the major materials' spectral properties, of which the object is composed. This knowledge may be obtained from a single calibration scan [4].

Section II outlines the theoretical aspects of the proposed algorithm. The method has been evaluated for simulated X-ray CT data, which is presented in section IV. Finally, we discuss relevant issues to conclude the paper in section V.

II. Methods

Assume a dataset consists of M materials with attenuation coefficients $\mu_m(E)$, $m = 1, ..., M$, which depend on the X-ray photon energy E. Specifying a reference density of each material, ρ_m, the mass attenuation coefficient for material m is $\kappa_m(E) = \mu_m(E)/\rho_m$. Denote $\vec{r}$ as spatial location on reconstruction grid. Knowing the spatial density distributions $\rho_m(\vec{r})$ and effective energy E_0, the monochromatic CT attenuation image can be calculated:

$$f(\vec{r}) = \sum_{m=1}^{M} \rho_m(\vec{r})\kappa_m(E_0). \tag{1}$$

$p_m(L) = \int_L dl\rho(\vec{r})$ is the line integral over projection ray L through a material density image $\rho(\vec{r})$. When a monochromatic X-ray beam traverses a homogeneous object, according to Lambert-Beer's law, the total attenuation coefficient is linearly related to the thickness of the object along the ray. The monochromatic intensity for a given E_0 can be expressed as

$$I_{\text{mono}}(L) = I_0(E_0) \cdot e^{-\sum_{m=1}^{M} p_m(L)\kappa_m(E_0)}. \tag{2}$$

However, in real CT, the emitted X-ray photons have varying energies $E \in [0, E_{max}]$. Therefore, the measured intensity of a polychromatic beam can be written as the sum of the monochromatic contributions for each energy E:

$$I_{\text{poly}}(L) = \int dE I_0(E) e^{-\sum_{m=1}^{M} p_m(L)\kappa_m(E)}, \tag{3}$$

where $I_0(E)$ is referred as normalized source-detector energy spectrum ($\int dE I_0(E) = 1$). The measured initial polychromatic attenuation $q_{\text{initial}}(L)$ along a ray path L is defined by

$$q_{\text{initial}}(L) = -\ln I_{\text{poly}}(L). \tag{4}$$

A. Single-material precorrection

Generally, when datasets consist of only one material, a precorrection can be carried out to reduce cupping artifacts caused by beam hardening. In clinical CT, water precorrection is widely used to reduce cupping artifacts. In industrial CT we usually precorrect for the most dominant material (i.e. the material which covers the most volume). The single material precorrection linearizes the projections at the first place in order to deliver improved and quantitative reconstruction for a better initial segmentation. In any case, the single material precorrection is a nonlinear preprocessing step of the initial rawdata:

$$q^{\text{precorr}}_{\text{initial}}(L) = q^{\text{precorr}}_{\text{initial}}(q_{\text{initial}}(L)). \tag{5}$$

B. Multi-material correction approach

A flowchart of the proposed algorithm is illustrated in Fig. 1. From the precorrected rawdata, we perform a preliminary filtered backprojection

$$g_{\text{initial}}(\vec{r}) = \mathcal{R}^{-1}\{q^{\text{precorr}}_{\text{initial}}(L)\}, \tag{6}$$

where $\mathcal{R}^{-1}$ denotes a filtered backprojection (FBP) reconstruction. The initial reconstruction is then segmented into M materials. We thereby require to know the number M of significant materials and the mass attenuation coefficient of these materials. The spectral properties can either be calibrated [4] or obtained as tabulated data from reference [15]. As quantitative CT values in the image are unreliable due to beam hardening, we use automatic centroids selection for k-means clustering [16], [17]. After segmenting $g_{\text{initial}}(\vec{r})$ into M masks $w_m(\vec{r})$, we maintain the original CT value by storing density volumes

$$\hat{\rho}_m(\vec{r}) = \frac{w_m(\vec{r}) \cdot g_{\text{initial}}(\vec{r})}{\kappa_m(E_0)}, \tag{7}$$

rather than the mask volumes $w_m(\vec{r})$. From the selected centroids at k-means clustering, effective energy E_0 can be obtained by choosing corresponding monochromatic energy at the centroid attenuation coefficients for each material. From the segmentation result, the CT attenuation image $g_{\text{initial}}(\vec{r})$ could be calculated according to Eq. (1). For each material, line integrals

$$\hat{p}_m(L) = \mathcal{R}\{\hat{\rho}_m(\vec{r})\} = \int_L dl\hat{\rho}_m(\vec{r}) \tag{8}$$

are calculated, where $\mathcal{R}$ denotes the calculation of line integrals through the volume along the originally measured lines L (forward projection). The line integrals are combined to a polychromatic rawdata set

$$\hat{q}(L) = -\ln \int dE I_0(E) e^{-\sum_{m=1}^{M} \hat{p}_m(L)\kappa_m(E)}, \tag{9}$$

which incorporates the spectral properties of each material $\kappa_m(E)$, as mentioned above. The polychromatic rawdata are then reconstructed again (including the single material precorrection) to obtain a recalculated image

$$g_{\text{calc}}(\vec{r}) = \mathcal{R}^{-1}\{\hat{q}^{\text{precorr}}(L)\}. \tag{10}$$

During segmentation and forward projection steps, errors which arise from beam hardening are additionally introduced.

The difference between the initially reconstructed volume and the recalculated volume can be used to estimate the beam hardening error:

$$g_{\text{BH}}(\vec{r}) = g_{\text{calc}}(\vec{r}) - g_{\text{initial}}(\vec{r}). \tag{11}$$

It has to be noted that spatial resolution mismatch occurs between $g_{\text{initial}}(\vec{r})$ and $g_{\text{calc}}(\vec{r})$. Therefore, a spatial resolution matching technique should be applied before the subtraction to maintain the spatial resolution of the final image. In general, resolution can be modulated by using boosting or smoothing kernels. In this paper, an optimized Gaussian smoothing kernel is applied to $g_{\text{initial}}(\vec{r})$ before the subtraction, such that

$$\hat{g}_{\text{BH}}(\vec{r}) = g_{\text{calc}}(\vec{r}) - \text{Gauss}(0,\sigma) * g_{\text{initial}}(\vec{r}), \tag{12}$$

with $\arg\min_\sigma ||\hat{g}_{\text{BH}}(\vec{r})||_2$. We finally use the spatially resolution-matched beam hardening image $\hat{g}_{\text{BH}}(\vec{r})$ to subtract the beam hardening from the initial reconstruction.

$$g_{\text{corrected}}(\vec{r}) = g_{\text{initial}}(\vec{r}) - \hat{g}_{\text{BH}}(\vec{r}). \tag{13}$$

As the corrected image could be used to obtain a better segmentation in the first place, the method can be repeated iteratively. Using the superset $i \geq 0$ to denote the iteration number and initializing with $g^0_{\text{corrected}}(\vec{r}) = g_{\text{initial}}(\vec{r})$, we can formulate a fixed-point equation

$$g^{i+1}_{\text{corrected}}(\vec{r}) = g_{\text{initial}}(\vec{r}) - \hat{g}_{\text{BH}}(g^i_{\text{corrected}}(\vec{r})). \tag{14}$$

III. Experiments

To evaluate the proposed algorithm, polychromatic cone-beam CT simulations were carried out using a FORBILD hip prosthesis phantom [18]. The dataset consists of soft tissue, bone, and the prosthesis (Ti). The projection data were obtained by using CT simulation software DRASIM (Siemens AG, Forcheim, Germany), and circular 3D raw data were reconstructed using a standard FDK reconstruction algorithm [19]. At a tube voltage of 100 kV we simulated 450 angular samples on a full circle, with a detector of 512×512 pixels, 0.5mm in pixel size. All projection images are reconstructed on a $512 \times 512 \times 512$ grid with a voxel size of 0.4mm.

For further examination of the algorithm, a real multi-material dataset containing four cylinders of different materials was evaluated. A 120 kV tube voltage was applied. Detector pixels are 1024×1024 with size of 0.4mm. All projection images are reconstructed on a $400 \times 400 \times 600$ grid with a voxel size of 0.5mm.

IV. Results and discussion

A. Beam hardening reduction

Fig. 2 and Fig. 4 illustrate the results from reconstructions and horizontal line profiles for simulated and real datasets, respectively. In comparison with original reconstructions (left), the recalculated polychromatic (middle) images show enhanced beam hardening artifacts. This illustrates our assumption in Eq. (12). The right figures show the final result, where beam hardening artifacts are hardly noticeable and the spatial resolution is maintained.

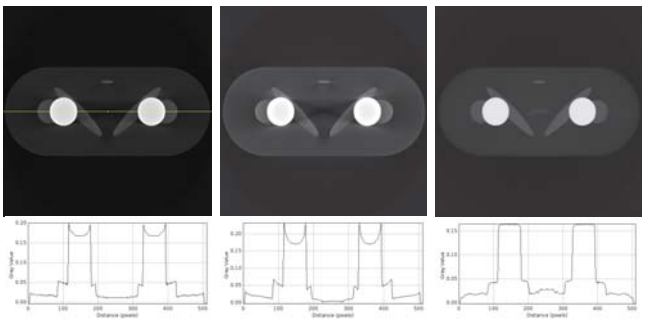


Fig. 2: Reconstruction results from hip prosthesis phantom with corresponding horizontal line profiles (yellow line). From left to right: original reconstruction, recalculated reconstruction, result after first iteration. The line profiles and images show attenuation coefficients (Level 0.10; window 0.22.)

Fig. 3: The beam hardening image $\hat{g}_{\text{BH}}(\vec{r})$ according to Eq. 12.

B. Spatial resolution matching

In Fig. 3 the beam hardening image $\hat{g}_{\text{BH}}(\vec{r})$ according to equation 12 is presented. The enhancement of object borders is visible from this image remains after spatial resolution matching and is caused by beam hardening.

Running a forward projection with subsequent reconstruction reduces the spatial resolution of an image. Calculating the difference between original and recalculated images would correspond to a high-pass filtering of the original image. As the initial image is linearly combined with the difference image, we need measurements to maintain the spatial resolution, especially when more than one iteration is applied.

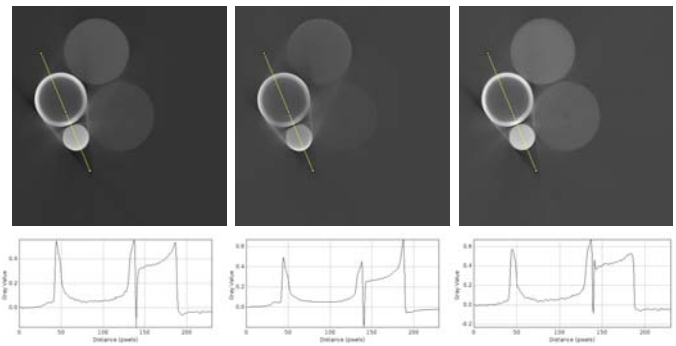
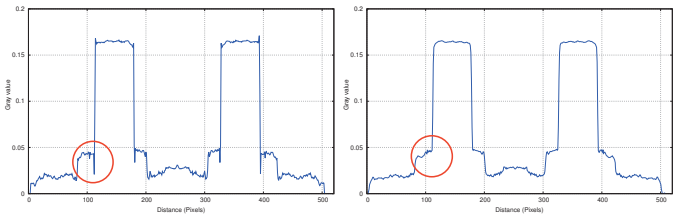
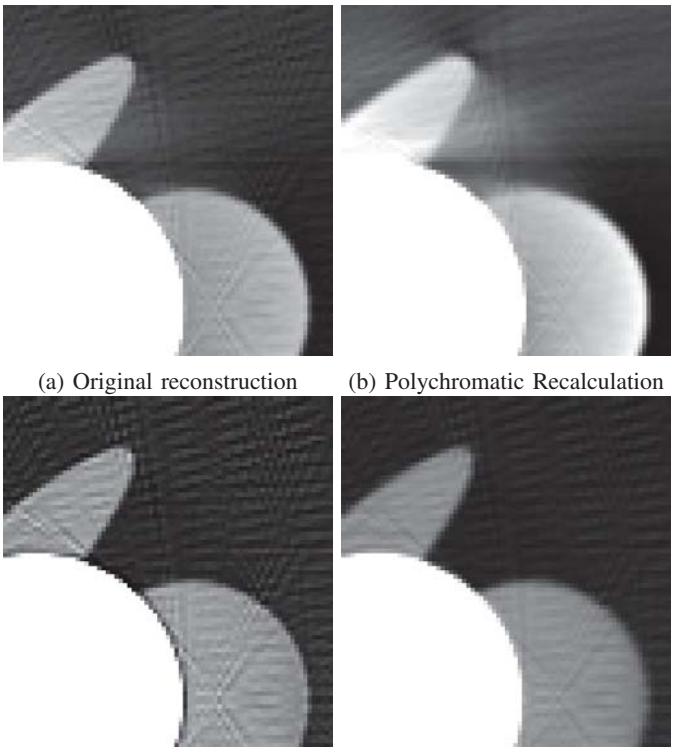


Fig. 4: Reconstruction results from 4-cylinder real dataset with corresponding horizontal line profiles (yellow line). From left to right: original reconstruction, recalculated reconstruction, result after first iteration. The line profiles and images show attenuation coefficients.



(a) (b)

Fig. 5: Horizontal line profile plots from corrected reconstructions with (a) and without (b) spatial resolution matching.



(a) Original reconstruction (b) Polychromatic Recalculation

(c) Correction without spatial resolution matching (d) Correction with spatial resolution matching

Fig. 6: Examination of spatial resolution matching using a zoom-in on the hip prosthesis phantom (Grayscale: level 0.04; window 0.06).

Fig. 5 presents the line profiles from reconstructions with (Fig. 5a) and without (Fig. 5b) the spatial resolution matching. It can be seen that calculation without consideration of resolution matching yields errors at object edges (red circle) and increases the level of aliasing.

A closer evaluation of spatial resolution influences on reconstructions of the hip prosthesis phantom are illustrated in Fig. 6. The polychromatic recalculation (Fig. 6b) has a lower resolution than the original reconstruction (Fig. 6a). Additionally, the enhanced cupping and streak artifacts can be noticed. Running the proposed method without spatial resolution matching, the object edges appear over-enhanced and the aliasing is increased as shown in Fig. 6c. However, if spatial resolution matching kernels are applied, beam hardening reduced reconstructions with maintained image impression can be achieved (Fig. 6d).

V. Conclusion

In this paper, an image-based beam hardening reduction algorithm that combines material density distribution with a polychromatic model of X-ray propagation is introduced. The algorithm has been implemented for a 3D cone beam geometry and was shown to yield excellent results in reducing cupping and streak artifacts. During iterations, segmentations with density information is retained to achieve more accurate results for reproducing a polychromatic model in forward projecting. In contrast to other iterative BHC approaches, our method distinctly preserves better original reconstruction information, which facilitates faster convergence. A spatial resolution matching technique is applied in order to improve image quality and overall performance.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] M. Kachelrieß, K. Sourbelle, and W. A. Kalender, "Empirical cupping correction: a first-order raw data precorrection for cone-beam computed tomography." *Med Phys*, vol. 33, no. 5, pp. 1269–1274, May 2006.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine and Biology*, vol. 21, no. 5, p. 733, 1976.

[3] J. P. Stonestrom, R. E. Alvarez, and A. Macovski, "A framework for spectral artifact corrections in x-ray CT," *Biomedical Engineering, IEEE Transactions on*, vol. BME-28, no. 2, pp. 128 –141, Feb. 1981.

[4] P. Stenner, T. Berkus, and M. Kachelrieß, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Medical Physics*, vol. 34, no. 9, pp. 3630–3641, 2007.

[5] C. Maaß, E. Meyer, and M. Kachelrieß, "Exact dual energy material decomposition from inconsistent rays (MDIR)," *Medical Physics*, vol. 38, no. 2, pp. 691–700, 2011.

[6] G. Herman, "Correction for beam hardening in computed tomography," *Physics in Medicine and Biology*, vol. 24, p. 81, 1979.

[7] E. V. de Casteele, "Model-based approach for beam hardening correction and resolution measurements in microtomography," Ph.D. dissertation, University Antwerpen, 2004.

[8] J. Hsieh, R. C. Molthen, C. A. Dawson, and R. H. Johnson, "An iterative approach to the beam hardening correction in cone beam CT," *Medical Physics*, vol. 27, no. 1, pp. 23–29, 2000.

[9] B. de Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT," *IEEE Trans. Med. Imaging*, vol. 20, no. 10, pp. 999–1008, 2001.

[10] I. Elbakri and J. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation," *Physics in Medicine and Biology*, vol. 48, p. 2453, 2003.

[11] M. Kachelrieß and W. Kalender, "Improving PET/CT attenuation correction with iterative ct beam hardening correction," in *Nuclear Science Symposium Conference Record (NSS/MIC)*, vol. 4, 2005.

[12] H. Gao, L. Zhang, Z. Chen, Y. Xing, and S. Li, "Beam hardening correction for middle-energy industrial computerized tomography," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, p. 2796, 2006.

[13] M. Krumm, S. Kasperl, and M. Franz, "Reducing non-linear artifacts of multi-material objects in industrial 3d computed tomography," *NDT & E International*, vol. Vol 41, No 4, pp. 242–251, 2008.

[14] G. V. Gompel, K. V. Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts, "Iterative correction of beam hardening artifacts in CT," *Medical Physics*, vol. 38, no. S1, pp. S36–S49, 2011.

[15] "NIST, the national institute of standards and technology." http://physics.nist.gov/physrefdata/xraymasscoef/cover.html, 2011.

[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes – The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.

[17] M. Tian, Q. Yang, A. Maier, I. Schasiepen, N. Maass, and M. Elter, "An automatic histogram-based initializing algorithm for K-means clustering in CT," in *Proceedings des Workshops Bildverarbeitung für die Medizin 2013*, 2013, pp. 277–282.

[18] http://www.imp.uni-erlangen.de/phantoms/hip/hipphantom.html.

[19] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America*, vol. 1, pp. 612–619, 1984.

Sparse-view cone-beam CT reconstruction via previous normal dose scan induced BM3D-frame regularization method

Liu Yang, Pascal Getreuer, and Linghong Zhou

***Abstract*—Cone-beam computed tomography (CBCT) images are obtained repeatedly in the course of radiation therapy. Scanning strategies such as the prior image constrained compressed sensing (PICCS) method of Chen et al. have been developed to take advantage of the redundancy between scans to enable reconstruction from fewer projections or lower dosage, thereby reducing the patient's radiation exposure. In this work, we use a prior full projection scan and the recent block matching 3D (BM3D) frame regularization of Danielyan et al. to guide the reconstruction from sparse-view data. The prior scan is first registered with an initial reconstruction estimate, and the BM3D frame is created. The prior image induced frame is then used to regularize the final reconstruction. Experiments show that this prior-BM3D-frame reconstruction strongly suppresses noise for sparse-view reconstruction compared with other methods. A fast algorithm is developed for solving the reconstruction problem using the split Bregman algorithm. This work may have great significance for image-guided radiation therapy when CBCT is used repeatedly.**

***Index Terms*— CBCT, Image Reconstruction, BM3D-frame, prior image constrains, split-Bregman**

I. Introduction

Cone-beam computed tomography (CBCT) technologies [1, 2] are widely adopted as an effective imaging tool in radiotherapy to help patient position setup and as an assistance in image guided radiation therapy (IGRT) [3, 4]. CBCT images are taken repeatedly during 2~4 weeks in one course of radiation therapy treatment. Reducing patients' accumulated radiation exposure is of top priority, as daily use of CBCT imaging produces a considerable amount of excessive radiation dose to radiotherapy patients [5, 6, 7]. However, conventional linear reconstruction algorithms such as Feldkamp–Davis–Kress (FDK) are significantly affected by geometric distortion and high noise ratios in low-dose scan protocols (such as fewer projections, lowering the tube current or voltage), leading to images of poor quality. Thus there is a demand for more robust reconstruction algorithms suitable for low dose scans.

Full projections are taken on the first day of radiation therapy treatment, and then in subsequent scans, the prior full projections are used as a reference, enabling reconstruction from only a sparse set of projections. Alternatively or in combination, subsequent scans can be made with lower dose in exchange for higher noise level. This prior image constrained scanning protocol shows a factor of 10–40 dose reduction compared to using full projections for all scans. We apply this scanning protocol in this work.

In this article, we propose a prior image induced 3D block matching frame regularization method (prior-BM3D-frame) for the sparse-projection reconstruction under the prior image constrained scanning protocol. The prior-BM3D-frame is motivated by the recent work on the frame interpretation of BM3D introduced by Danielyan, Katkovnik, and Egiazarian [8] and the prior image constrained compressed sensing (PICCS) method of Chen et al [9].

The innovation of our proposed method lies in that unlike [8], which generate the BM3D-frame transformation matrices according to the initial estimated image, we construct the BM3D-frame transformation matrices from a registered image between the prior image and roughly reconstructed sparse-view image using conventional linear reconstruction algorithms. The method is illustrated in Section II and the effectiveness of our proposed prior-BM3D-frame approach is shown numerically in Section III.

II. Methodology

A. Model

The CBCT image reconstruction problem can be treated as a discrete linear inverse problem:

$$Au = b, \tag{1}$$

where u is a N -dimensional vector of attenuate coefficients for reconstruction and b is a M -dimensional vector of log-transformed data for each projection, A is a $M \times N$ system matrix that includes weighting values for each pixel, and could be generated by Siddon's ray-tracing algorithm [10] after the projection geometries for each patient case is given.

As conventional linear reconstruction algorithms (i.e. FBP for 2-dimensional case and FDK for 3D case) for solving the

This work was partially supported by the National Natural Science Foundation of China (Grant No. 30970866) and Guangdong Strategic Emerging Industry Core Technology Research (Grant No.2011A081402003).

L. Yang and L. Zhou are with the School of Biomedical Engineering, Southern Medical University, Guangzhou, GD 510515 P.R.China (e-mail: snoopydowell@gmail.com; smart@smu.edu.cn).

P. Getreuer is with the CMLA lab, ENS Cachan, 61, avenue du Président Wilson, 94235 Cachan cedex, France. (e-mail: getreuer@cmla.ens-cachan.fr).

under-determined equation may not be feasible due to the limited number of views and the noise in the projection data, more decent and effective methods are called for. The emergence of Compressed Sensing (CS) theory [11] provides a new prospective for solving these kind of problems by minimizing the following objective function iteratively:

$$\min_u \frac{\mu}{2} \| Au - b \|_2^2 + \lambda | \Phi u |_1 \tag{2}$$

The first term is for data-fidelity where $\| \cdot \|_2^2$ denotes the Euclidean norm, the second term plays as a penalized regularization term where $| \cdot |_1$ denotes the ℓ^1 norm, $\mu > 0$ and $\lambda > 0$ are weighting parameters that decides the proportion of data-fidelity and regularization in equation respectively. Φ is defined as sparse representation transformation.

Danielyan et al. [8] showed that the BM3D filtering technique can be formulated in terms of an overcomplete sparse frame representation to extend the scope of BM3D filtering to deblurring and other inverse problems. We present here the analysis based prior image induced BM3D-frame for sparse-view CBCT image reconstruction. The analysis frame interpretation of BM3D is to reconstruct the image u by solving the convex minimization problem (2), and Φ is block-matching 3D frame transformation matrix which acts as the *Patch matching and Analysis* substeps of BM3D filtering method [12, 13].

We exploit the prior image u_{pri} generated from full-scan as a patch-matching reference, we use the registered image u_{reg} acquired from the demons registration [14]

$$u_{reg} = \frac{(u_{pri} - u_{est}) \nabla u_{est}}{| \nabla u_{est} |^2 + (u_{pri} - u_{est})^2} \tag{3}$$

where u_{est} is the roughly reconstructed image generated by sparse-view scan using FBP as the static image. The iteration times of demons algorithm is set to 5 according to the pre-tests within the clinical position setup error [15].

The patch matching substep is defined as followes in detail: Let y denote the noisy image and let y_m denote the $\kappa \times \kappa$ square patch of y whose top-left corner is pixel m of the image. Let P_m denote the patch extraction operator such that

$$P_m y = y_m \tag{4}$$

For each y_m , the image is searched over a square neighborhood with side N_s for similar patches,

$$N(m) = \{ n : \| y_m - y_n \|_2 \le \tau \} \tag{5}$$

where τ is a threshold. 3D groups are formed by stacking similar patches. The m th 3D group is:

$$\underset{n \in N(m)}{\text{stack}} \, y_n \tag{6}$$

where stack denotes stacking of $\kappa \times \kappa$ patches to create a 3D array of size $\kappa \times \kappa \times | N(m) |$. Since $N(m)$ is a set, its order is arbitrary, we put the set in a declination order according to the value of $\| y_m - y_n \|_2$.

For the m th 3D group, define Φ_m as:

$$\Phi_m u_{reg} \coloneqq \Gamma_{3D} \underset{n \in N(m)}{\text{stack}} P_n u_{reg} \tag{7}$$

where u_{reg} denotes a $\sqrt{N} \times \sqrt{N}$ square matrix of attenuate coefficients acquired from the Demons registration, P_n denotes the patch extracts operators defined in (4). The 3D transform Γ_{3D} is constructed as a separable combination of a 2D transform Γ_{2D} applied on each patch and a 1D transform Γ_{1D} applied along the third dimension of the 3D group. This structure allows Γ_{2D} to be applied to each patch before stacking,

$$\Gamma_{1D} \underset{n \in N(m)}{\text{stack}} \Gamma_{2D} y_n \tag{8}$$

Reasonable choices for the 1D and 2D transforms include wavelet transforms, discrete cosine and sine transforms, and the Walsh–Hadamard transform. We use a 1D Walsh–Hadamard transform for Γ_{1D} and the "bior1.5" biorthogonal spline wavelet transform with 1 analysis and 5 synthesis vanishing moments for Γ_{2D} as suggested by [13].

B. Algorithm

The formation of Φ in (2) can be concluded as: Φ_m extracts patches, stacks them into a 3D group of size $\kappa \times \kappa \times | N(m) |$, and applies the transform Γ_{3D}. In matrix-vector notation, $u \in R^N$ where N is the number of pixels in the image and Φ_m is a $\kappa \times \kappa \times | N(m) |$ by N matrix. Φ_m are vertically concatenated to create matrix Φ of size $\kappa \times \kappa \times | N(m) |$ by N ,

$$\Phi u = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_M \end{bmatrix} u. \tag{9}$$

For the solution of (7), it will be useful to compute the transpose of Φ^T . It can be represented functionally as

$$\Phi^T \omega = \sum_m \sum_{n \in N(m)} P_n^T (\Gamma_{3D}^T \omega_m)_n \tag{10}$$

where ω_m is the $\kappa \times \kappa \times | N(m) |$ array of coefficients corresponding to the m th 3D group of ω and $(\Gamma_{3D}^T \omega_m)_n$ denotes the nth patch in the 3D group.

With all the parameters required (system matrix A , projection data b , registered image u_{reg} , sparse transformation matrix Φ , and initial estimated image u_{est} reconstructed by FBP using sparse view projection data) being calculated or detected, we use the split Bregman method proposed by Goldstein and Osher [16] for solving equation (2),

which has been proved efficiently in many related applications [17, 18]. The whole algorithm can be concluded as:

Input: projection data b, system matrix A, initial estimated image u_{est} using FBP, prior full-scan image u_{pri},

1: Demons registration: $u_{reg} = \text{Demons}\,(u_{pri}; u_{est})$,

2: calculate BM3D frame matrices using equation (9) and (10): Φ, Φ^T

3: Initial estimate: $u_0 = u_{est}$; $d_0 = 0$; $f_0 = 0$;

4: **while** $|u^{k+1} - u^k| \geq 1\times10^{-4}$ **do**

5: (a) CGLS (solve for u^{k+1}):
$\mu A^T A u^{k+1} = \mu A^T b + \lambda \Phi^T (d^k - \Phi u^{k+1} - f^k)$

6: (b) Shrinkage:
$d^{k+1} = \text{shrink}(\Phi u^{k+1} + f^k, 1/\lambda)$
where $\text{shrink}(x) = \text{sign}(x)\max\{|x| - \frac{1}{\lambda}, 0\}, \quad x \in R$

7: (c) Update f :
$f^{k+1} = f^k + \Phi u^{k+1} - d^{k+1}$

8: $k = k+1$

9: **end while**

Output: reconstructed final image u.

The parameters μ and λ are empirically chosen and the optimized μ and λ are case-to-case different. However, according to the pretests, our method performs better with parameter $\mu = 100 \sim$? 00 while $\lambda = 0.1 \sim$? times that of μ with different number of views for our experiment cases. The stopping criteria $|u^{k+1} - u^k|$ is set to 1×10^{-4} for a balance between better image quality and algorithm speed.

III. Results

In this proof-of-concept study, we implement our proposed prior-BM3D-frame algorithm in 2D phantom, rather than 3D, in a 512 by 512 spatial grid. Projection data were acquired by Siddon's algorithm in the 2D fan beam scanning geometry. The distance from the source to the isocenter is 100cm and the distance from source to the detector center is 150cm., being similar to a typical middle slice of a cone-beam CT. We implement the sparse-view protocol with different number of projections: $N_\theta = 10, 20, 40.$

A. Phantom and Geometry Setup

The Simulated digital NURBS-based cardiac-torso (NCAT) phantom Figure 1(a), which maintains a high level of detailed anatomical realism Figure 1(b), is generated at one thorax region slice of size 512×512 pixels, and the X-ray detector is modeled as a 512-element vector . In the modeled sparse-view protocol, the projections are generated on equally spaced views covering an entire 360° rotation with Siddon's ray-tracing algorithm [10], performed on the NCAT phantom (Figure 1(a)). The prior image we use here is the NCAT phantom with slight rotation (2^o) and translation (10 pixels) that is set to simulate the clinical position setup error [15]. The NCAT phantom (Figure 1(a)) serves as the ground truth, the rotated and translated NCAT phantom shows the prior image.

B. Prior-BM3D-frame implementation

We execute our algorithm with parameters selected as in Section II, with $\lambda = 0.2\mu$ while $\mu = 1000$ for $N_\theta = 20, 40$ and $\mu = 100$ for $N_\theta = 10$. We also compare our proposed algorithm with conventional FBP method with Hamming filter, TV minimization and the BM3D-frame regularization without using a prior image. By omitting the substeps 1 and 2 in the left column and calculating Φ, Φ^T using u_{est} instead of u_{reg}, we can easily get BM3D-frame regularization without prior image. The TV minimization formula is as follows:

$$\min_u \frac{\mu}{2} \| Au - b \|_2^2 + \lambda |\nabla u|_1 \tag{11}$$

Both TV minimization and BM3D-frame without prior image are solved iteratively by the split-Bregman method. For comparison, the parameters are set the same as in prior-BM3D-frame. With the priors, the streaking artifacts on the images are strongly suppressed compared to the FBP reconstructions, and the image is detailed in structure and small bronchioles are visually discernible as shown in Figure 2, 3, 4. It is easy to observe that 20 projections are enough for prior-BM3D-frame to reconstruct image with clear details, while FBP, TV and BM3D-frame cannot .

C. Quantify the results

We quantify the results by computing the relative error (Err) defined as

$$Err = \frac{\| u - u_0 \|_2}{\| u_0 \|_2} \tag{12}$$

where u_0 is the ground truth and u is the reconstructed image. See Table I for the quantitative summarization. Figure 5 shows the faster convergence of prior-BM3D-frame.

(a) (b)

Fig.1. The NCAT phantom generated in a 512 by 512 spatial grid. (a) the Ground Truth, and (b) the prior image (the NCAT phantom with slight rotation (2^o) and translation (10 pixels)

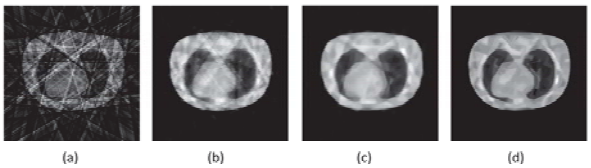


Fig. 2. Comparison of the images reconstructed by FBP, TV, BM3D-frame, a priori-BM3D-frame with 10 projections generated by NCAT phantom. The window length for each image is adjusted to [0 MAX], where MAX represents the largest.

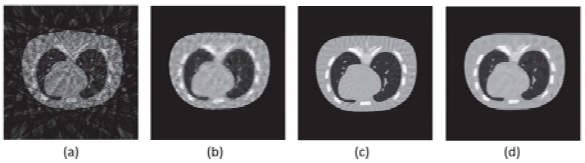


Fig. 3. Comparison of the images reconstructed by FBP, TV, BM3D-frame, a priori-BM3D-frame with 20 projections generated by NCAT phantom. The window length for each image is adjusted to [0 MAX], where MAX represents the largest.

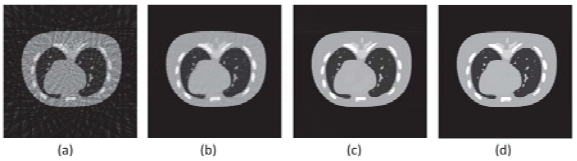


Fig. 4. Comparison of the images reconstructed by FBP, TV, BM3D-frame, a priori-BM3D-frame with 40 projections generated by NCAT phantom. The window length for each image is adjusted to [0 MAX], where MAX represents the largest.

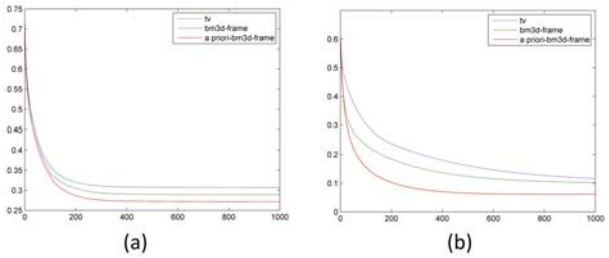


Fig. 5. The decline curve of relative error using 10 and 20 projections of NCAT phantom. (a) TV, BM3D-frame, a priori-BM3D-frame convergence curve using 10 projections. (b) TV, BM3D-frame, a priori-BM3D-frame convergence curve using 20 projections.

TABLE I
RELATIVE ERROR BETWEEN THE RECONSTRUCTED IMAGES USING FBP, TV, BM3D-FRAME, PRIOR-BM3D-FRAME AND GROUND TRUTH.

# of projection	FBP	TV	BM3D-Frame	prior-BM3D-Frame
10	0.7606	0.3086	0.2851	0.2714
20	0.5322	0.1253	0.1037	0.0627
40	0.3143	0.0568	0.0427	0.0098

ACKNOWLEDGMENT

The authors thank Dr. Hao Gao (Math&CS Dept, Emory University) for valuable discussions of this manuscript.

REFERENCES

[1] D. A. Jaffray, D. Drake, M. Moreau, A. A. Martinez, J. W. Wong, A radiographic and tomographic imaging system integrated into a medical linear accelerator for localization of bone and soft-tissue targets, International Journal of Radiation Oncology, Biology, Physics 45 (1999) 773–789.

[2] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, A. A. Martinez, Flat-panel cone-beam computed tomography for image-guided radiation therapy, International Journal of Radiation Oncology, Biology, Physics 53 (2002) 1337–1349.

[3] C. A. McBain, A. M. Henry, J. Sykes, A. Amer, T. Marchant, C. M. Moore, J. Davies, J. Stratford, C. McCarthy, B. Porritt, P. Williams, V. S. Khoo, P. Price, X-ray volumetric imaging in image-guided radiotherapy: The new standard in on-treatment imaging, International Journal of Radiation Oncology, Biology, Physics 64 (2006) 625–634.

[4] I. S. Grills, G. Hugo, L. L. Kestin, A. P. Galerani, K. K. Chao, J. Wloch, D. Yan, Image-guided radiotherapy via daily online cone-beam CT substantially reduces margin requirements for stereotactic lung radiotherapy, International Journal of Radiation Oncology, Biology, Physics 70 (2007) 1045–1056.

[5] M. K. Islam, T. G. Purdie, B. D. Norrlinger, H. Alasti, D. J. Moseley, M. B. Sharpe, J. H. Siewerdsen, D. A. Jaffray, Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy, Med. Phys. 45 (2006) 773C–89.

[6] M. Daly, J. Siewerdsen, D. Moseley, D. Jaffray, J. Irish, Intraoperative cone-beam ct for guidance of head and neck surgery: assessment of dose and image quality using a c-arm prototype, Med. Phys. 33 (2006) 3767.

[7] G. X. Ding, C. W. Coffey, adiation dose from kilovoltage cone beam computed tomography in an image-guided radiotherapy procedure, Int. J. Radiat. Oncol. Biol. Phys. 73 (2009) 610–7.

[8] A. Danielyan, V. Katkovnik, K. Egiazarian, BM3D frames and variational image deblurring, IEEE Transactions on Image Processing 21 (2012) 1715–1728.

[9] G.-H. Chen, J. Tang, S. Leng, Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets, Medical Physics 35 (2008) 660–663.

[10] R. L. Siddon, Fast calculation of the exact radiological path for a three-dimensional CT array, Medical Physics 12 (1985) 252–255.

[11] D. L. Donoho, "Compressed sensing", *IEEE Trans. Inf. Theory*, 52, 1289-306 (2006).

[12] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising with block-matching and 3D filtering, volume 6064 of Proceedings of SPIE, SPIE, 2006.

[13] M. Lebrun, An analysis and implementation of the BM3D image denoising method, 2012. Preprint.

[14] J.-P. Thirion, Image matching as a diffusion process: an analogy with Maxwell's demons, Medical Image Analysis 2 (1998) 243–260.

[15] C. Thilmann, S. Nill, T. Tucking, A. Hoss, B. Hesse, L. Dietrich, R. Bendl, B. Rhein, P. Haring, C. Thieke, U. Oelfke, J. Debus, P. Huber, Correction of patient positioning errors based on in-line cone beam cts: clinical implementation and first experiences, Radiation Oncology 1 (2006) 16.

[16] T. Goldstein, S. J. Osher, The split bregman method for l1 regularized problems, SIAM Journal on Imaging Sciences 2 (2009) 323–343.

[17] P. Getreuer, Rudin–Osher–Fatemi total variation denoising using split Bregman, Image Processing On Line (2012). DOI: http://dx.doi.org/10.5201/ipol.2012.g-tvd.

[18] H. Gao, J. F. Cai, Z. Shen and H. Zhao, "Robust principle component analysis based four-dimensional computed tomography". *UCLA CAM Report*, 10-79 (2010).

[19] W. Segars, B. Tsui, D. Lalush, E. Frey, M. King, D. Manocha, Development and application of the new dynamic.

[20] NURBS-based cardiac-torso (ncat) phantom, Biomedical 5 Engineering, (Chapel Hill, NC: University of North Carolina) (2011).

An algorithm for full 3D reconstruction with an arbitrary trajectory

Victor P. Palamodov *Fellow, AMS*

Abstract-**A method of full 3D reconstruction from X-ray data is presented for a source orbit of general form. The algorithm is of $O\left(N^4\right)$ operations long**.[1]

I. Analytic reconstruction

Given a function f in an Euclidean 3D space E with compact support, the Fourier transform

$$F(f)(\xi)=\int_E f\exp\left(-2\pi i\left\langle \xi,x\right\rangle\right)\mathrm{d}V$$

can be recovered from sufficiently rich data of ray integrals

$$g(x,\nu)=\int_0^\infty f(x+tv)\,\mathrm{d}t$$

where x is a source point and v is a unit direction vector, A family Σ of rays $R\subset E^3$ is called *complete* in a set $U\subset E^3$ if for any point $p\in U$ and any plane H through p there exists a ray $R\in\Sigma$ such that $p\in R\subset H$.

Theorem. *Let Γ be a piecewise C^1-continuous curve in E^3 such that the family $\Sigma(\Gamma)$ of rays with sources in Γ is complete in a compact set $K\subset E$. For any function f with support in K the Fourier transform $F(f)$ can be reconstructed as follows*

$$F(f)(\tau\omega)=\frac{1}{i\tau}\int_{\mathbb{R}}\exp\left(-i2\pi\tau p\right)\mathrm{d}p \cdot\sum\varepsilon(s,\omega)\int_{v\in\mathrm{C}_\omega}\left\langle\omega,\nabla_v\right\rangle g\left(x(s),v\right)\mathrm{d}v \quad (1)$$

Here ω is an arbitrary unit vector and $\tau\in\mathbb{R}$,
$x=x(s)$, $s\in[0,1]$ is a parametrization of Γ and $p=\left\langle x(s),\omega\right\rangle$, $\mathrm{d}p=\left\langle x'(s),\omega\right\rangle ds$
C_ω is the unit circle in a plane orthogonal to ω and $\mathrm{d}v$ is the angular measure on the circle,
ε is an arbitrary bounded function in $[0,1]$ satisfying a normalizing condition

$$\sum_{x(s)\in H(\omega)}\varepsilon(s,\omega)=1$$

for any plane $H(\omega)$ orthogonal to ω that meet the support of f.

A proof is straight forward. Existence of a function ε follows from the completeness assumption.

II. Description of algorithm

An algorithm based on () can be implemented in a few steps:

1. Calculate of a function ε. A continuous (e. g. a piecewise linear) function ε is preferable to suppress artefacts.

2. Scale the Euclidean metric in such a way that $\mathrm{diam}K=1$. Let N be the radius of a frequency domain where we want to obtain reliable resolution. Choose a stepsize $h\approx\sqrt{3}N$ and take a set $\Omega\subset E$ of $O\left(N^2\right)$ such that $\mathrm{dist}(\theta,\Omega)\leq h$ for an arbitrary unit vector θ.

3. For each $\omega\in\Omega$ we choose a grid G_ω on the line $\mathbb{R}\omega$ with stepsize h. Take all points $s=s_1,...,s_M\in[0,1]$ such that $\varepsilon(s_i,\omega)\neq 0$ and $p=\left\langle x(s_i),\omega\right\rangle$ are nodes of the grid. We may assume that $M\leq 2N$.

4. For any $\omega\in\Omega$ and any $i=1,...,M$ determine a set of detectors $d_k,k=1,...,O(N)$ that are nearest to the plane $H(p,\omega)$ through $x(s_i)$ orthogonal to ω.

5. Calculate the derivative $h(\omega,v)=\left\langle\omega,\nabla_v\right\rangle g\left(x(s_i),v\right)$ for

$$v=v_{ik}=\frac{d_k-x(s_i)}{|d_k-x(s_i)|},\ k=1,...,O(N)$$

An interpolation can be used in the set of detectors close to $H(\omega)$.

6. Calculate the sums

$$S(p,\omega)\doteq\sum_{i,\left\langle x(s_i),\omega\right\rangle=p}\varepsilon(s_i,\omega)\sum_k h(\omega,v_{ik})\,\varphi_{ik}$$

for each point $p\in G_\omega$ where φ_{ik} is the angle between v_{ik} and v_{ik-1}. All the terms vanish if p does not belong to the projection of K to the line $\mathbb{R}\omega$ hence the support of S is contained in an interval of length 1. Steps 5 and 6 need $O\left(N^2\right)$ operations for each vector $\omega\in\Omega$ which costs totally $O\left(N^4\right)$ operations.

7. Last steps are similar to the modified Fourier algorithm for Radon transform as in [1] ,Ch.V. First, calculate the discrete Fourier transform $\hat{S}(\tau,\omega)=F_{p\mapsto\tau}S(p,\omega)$ for $\tau=1,...,T$. This consumes $O(N\log N)$ for each ω and $O\left(N^3\log N\right)$ operations for all ω.

8. Interpolate the function $F_S(f)(\tau\omega)=(i\tau)^{-1}\hat{S}(\tau,\omega)$ form the spherical grid $\{\zeta=\tau\omega,0<\tau\leq T,\omega\in\Omega\}$ to a function $F_C(f)$ on a Cartesian grid $C=\{k\in Z^n,|k_j|\leq N,j=1,...,n\}$ and set $F_Cf(0)=0$. Here we use an improved interpolation method as in [1].

9. Calculate the inverse Fourier transform $f_C=F^{-1}\left(F_C(f)\right)$. The last two steps may consume $O\left(N^3\log N\right)$ operations.

10. The reconstruction f_C is an approximation to $f+\mathbf{c}_0$ where $\mathbf{c}_0$ is constant due to vagueness of $F_Cf(0)$ (expected small). This constant can be easily found since $f=0$ on the boundary of K.

Thus the total number of operation is still $O\left(N^4\right)$ whereas the standard algorithms [2, 3, 4] cost apparently $O\left(N^5\right)$ operations and algorithm 1-10 might be competitive.

References

[1] F. Natterer. The mathematics of computerized tomography. B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester, 1986.

[2] A. Katsevich. An improved exact filtered backprojection algorithm for spiral computed tomography, *Advances in Applied Mathematics*, vol. 32, 2004, pp. 681-697.

[3] J. D. Pack and F. Noo. Cone-beam reconstruction using 1D filtering along the projection of M-lines, *Inverse Problems* vol. 21, 2005, pp. 1105–1120.

[1]Address: Tel Aviv University, Israel. Email: palamodo@post.tau.ac.il

[4] A. Katsevich and M. Kapralov. Filtered backprojection inversion of the cone beam transform for a general class of curves, *SIAM J. Appl. Math*. vol. 68, 2007, pp. 334–353.

Tomographic Neutron Imaging using SIRT

Jens Gregor
University of Tennessee
jgregor@eecs.utk.edu

Charles Finney
Oak Ridge National Laboratory
finneyc@ornl.gov

Todd Toops
Oak Ridge National Laboratory
toopstj@ornl.gov

***Abstract*— Neutron imaging is complementary to x-ray imaging in that materials such as water and plastic are highly attenuating while material such as metal is nearly transparent. We showcase tomographic imaging of a diesel particulate filter. Reconstruction is done using a modified version of SIRT called PSIRT. We expand on previous work and introduce Tikhonov regularization. We show that near-optimal relaxation can still be achieved. The algorithmic ideas apply to cone beam x-ray CT and other inverse problems.**

I. Introduction and Background

Neutrons can be used for tomographic imaging in a manner similar to x-rays since both are subject to absorption and scattering when penetrating matter which leads to contrast. The main difference between the two is that neutron attenuation depends on a material property called the neutron cross-section but not on the atomic number which influences the probability of x-ray attenuation. Neutrons are thus attenuated more so by hydrogenous materials such as water and plastic than by dense material such as metal. In contrast, x-rays are attenuated less so by the former than by the latter. As illustrated in Fig. 1, neutron and x-ray images therefore convey complementary information.

A main goal of neutron imaging is to determine the material level macroscopic cross-section $\Sigma = N\sigma$ where N denotes atomic density and σ denotes the atomic level microscopic cross-section which models the energy dependent likelihood that a neutron will interact with a target nucleus. From this, it can be seen that knowledge of Σ provides indirect information about other material properties since $\sigma = \rho N_A/M$ where ρ is material density, N_A is Avagadro's number, and M is molar mass. Macroscopic cross-section has dimension of 1/cm which makes it analogous to the linear attenuation coefficient of x-rays. Indeed, from a mathematical point of view, obtaining one is no different from obtaining the other since Beer's Law describes the behavior of both [1]. For neutrons:

$$I_1 = I_0 \; e^{-\int_L dl \; \Sigma} \tag{1}$$

where I_0 and I_1 denote the intensity of the incident and transmitted neutron beam, and L denotes the beam path through the object.

Neutron beams are typically generated by a reactor but could also be produced by a spallation source. The experimental work presented here is based on the former and showcases data from the High Flux Isotope Reactor (HFIR) located at Oak Ridge National Laboratory in Oak Ridge, Tennessee. Used for a wide variety of applications ranging from food science and plant systems biology to energy storage and material science, the CG-1D beamline was recently used to image diesel particulate filters which serve to prevent soot and other biologically and environmentally harmful particulates from being emitted into the air. We discuss the CG-1D beamline and the diesel particulate filter application in more detail below.

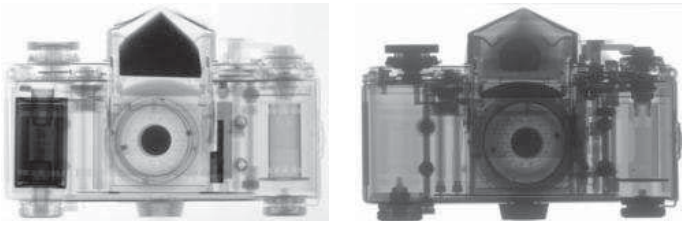


Fig. 1. Projection image of camera casing obtained using neutrons (left) and x-rays (right). [Images from http://www.psi.ch/niag/what-is-neutron-imaging]

Log-normalization and discretization of (1) yields a linear system of equations which we solve using the SIRT (Simultaneous Iterative Reconstruction Technique) algorithm [2]. Having been applied to a wide range of inverse problems in medicine and biology as well as other fields, SIRT is an example of a Richardson Iteration which is a classical numerical method for solving a linear system of equations. We have previously shown that near-optimal relaxation can be achieved by means of eigenvalue analysis and introduced a scalar-preconditioned version called PSIRT [3]. We improve on this work by introducing first a simple way to compute an upper bound on the smallest eigenvalue and then adding Tikhonov regularization in the form of a minimum norm constraint within the relaxed PSIRT framework. The algorithmic ideas are generic and apply to cone beam x-ray CT and other inverse problems.

II. SIRT: Algorithmic Development

Let x and b denote image and log-normalized projection data, A the system matrix that connects the two, and R and C two diagonal matrices of inverse row and column sums of A. Then SIRT solves the weighted least-squares problem:

$$\mathrm{x}^* = \operatorname{argmin} \|\mathrm{Ax} - \mathrm{b}\|^2_{\mathrm{R}} \tag{2}$$

The solution is computed using the relaxed iteration:

$$\mathrm{x}^{(k+1)} = \mathrm{x}^{(k)} + \alpha \; \mathrm{CA}^{\mathrm{T}}\mathrm{R}(\mathrm{b} - \mathrm{Ax}^{(k)}). \tag{3}$$

In practice, updating takes place using an ordered subsets approach. See [3], [4] for details.

A. Near-Optimal Relaxation

SIRT is a Richardson Iteration. Convergence is guaranteed if $0 < \alpha < 2/\lambda_{\max}$ with the fastest rate of convergence obtained for $\alpha^* = 2/(\lambda_{\max} + \lambda_{\min})$. Here $\lambda_{\max}$ and $\lambda_{\min}$ refer to the largest and smallest eigenvalues of matrix $\mathrm{CA^TRA}$. Assuming A has full column rank, both eigenvalues are strictly positive.

Stemming from the fact that non-negative matrices $\mathrm{CA^T}$ and RA are both stochastic, it follows that $\lambda_{\max} = 1$. This implies $1 \leq \alpha^* = 2/(1 + \lambda_{\min}) < 2$. Empirical comparisons of residual norms have consistently found $\alpha = 1.99$ to converge twice as fast as $\alpha = 1.00$ in terms of requiring half as many iterations to achieve the same result. This indicates $\lambda_{\min} \ll 1$. Until now this finding has not been quantified, but we here detail a straightforward pathway.

The trace of an N×N matrix is equal to the sum of the eigenvalues of that matrix [5]. Clearly, the smallest eigenvalue must be less than or equal to the average of all eigenvalues. For SIRT, we can establish an even tighter bound, namely:

$$\lambda_{\min} \leq \frac{\mathrm{tr}(\mathrm{CA^TRA}) - 1}{\mathrm{N} - 1} \tag{4}$$

where we have subtracted off the value of the largest eigenvalue and computed the average of the remaining eigenvalues. We will use $\lambda^*_{\min}$ to refer to this bound below.

B. Scalar Preconditioning

Matrix C merely serves to precondition the normal equations associated with (2). We have introduced an alternative scalar preconditioning scheme [3]. The resulting PSIRT algorithm is given by:

$$\mathrm{x}^{(k+1)} = \mathrm{x}^{(k)} + \alpha\,\mathrm{p}\,\mathrm{A^TR}(\mathrm{b} - \mathrm{Ax}^{(k)}) \tag{5a}$$

$$\mathrm{p} = 1/\|\mathrm{A}\|_1 \tag{5b}$$

where $\|\mathrm{A}\|_1 = \max_j \sum_i a_{ij}$ denotes the maximum column sum of the system matrix. The advantage of PSIRT over SIRT is best seen when ordered subsets are used and the code is executed in a distributed environment. SIRT requires matrix C to be recomputed each iteration on a per subset basis which is costly as it calls for a global reduction of equally many image-sized data structures. In contrast, PSIRT need do so only the first time as the set of scalars replacing the matrices can conveniently be stored for future use.

C. Tikhonov Regularization

The condition number of matrix A is the ratio of its largest and smallest singular values [5]. A small condition number indicates that the linear system $\mathrm{Ax} = \mathrm{b}$ can be solved with great precision. Conversely, a large condition number indicates that no algorithm can guarantee to find a solution with any provable accuracy.

Tikhonov regularization is a widely used technique for improving the numerical stability of an algorithm for solving a poorly conditioned linear system [6]. In our case, it leads to the weighted least-squares problem:

$$\mathrm{x}^* = \mathrm{argmin}\,\|\mathrm{Ax} - \mathrm{b}\|^2_{\mathrm{R}} + \beta\,\|\mathrm{Qx}\|^2 \tag{6}$$

where matrix Q is chosen to emphasize structural characteristics of x that are undesirable. We obtain a SIRT-like update scheme by preconditioning the normal equations associated with (6) by matrix C followed by matrix splitting:

$$\mathrm{x}^{(k+1)} = (\mathrm{I} - \beta\,\mathrm{CQ})\mathrm{x}^{(k)} + \alpha\,\mathrm{CA^TR}(\mathrm{b} - \mathrm{Ax}^{(k)}). \tag{7}$$

We consider the special case where $\mathrm{Q} = \mathrm{I}$ for which preference is given to a minimum norm solution. Combined with the scalar preconditioning used by PSIRT, we obtain the following simplified update scheme:

$$\mathrm{x}^{(k+1)} = (1 - \beta\,\mathrm{p})\mathrm{x}^{(k)} + \alpha\,\mathrm{p}\,\mathrm{A^TR}(\mathrm{b} - \mathrm{Ax}^{(k)}). \tag{8}$$

Near-optimal relaxation is achieved for $\alpha^* = 2/(1 + \lambda^*_{\min} + 2\,\beta\,\mathrm{p})$. Assuming $\lambda^*_{\min}$ is negligible, a value of $\beta\,\mathrm{p} < 0.025$ leads to $\alpha > 1.90$ which is within 5 percent of the 1.99 value achieved without regularization.

The condition number for the underlying linear system is given by:

$$\kappa = \sqrt{\frac{\sigma^2_{\max} + \beta\,\mathrm{p}}{\sigma^2_{\min} + \beta\,\mathrm{p}}} \tag{9}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ denote the largest and smallest singular values of matrix $\mathrm{pA^TRA}$. Regularization is needed when $\sigma_{\min}$ is close to zero. In this case, which follows when $\lambda^*_{\min}$ is close to zero since $\sigma_{\min} \leq \lambda_{\min}$, the approximation $\kappa = \sigma_{\max}/\sqrt{\beta\,\mathrm{p}}$ implies even a relatively small value of $\beta\,\mathrm{p}$ has the potential to greatly improve the conditioning and thus the numerical stability of a PSIRT based reconstruction. We have found $\beta = 0.01/\mathrm{p}$ to strike a good balance between relaxation based on $\alpha = 1.95$ and thus a less than two percent decrease in the maximum rate of convergence rate increase possible, and achieving regularization approximately equal to $\kappa = 10\sigma_{\max}$ which is a substantial improvement relative to the unregularized case.

III. SIRT: Implementation Overview

At Fully3D 2011, we presented an implementation of PSIRT for execution on a small cluster of multi-core PCs [4]. We briefly summarize that work as it forms the foundation for the regularized version of PSIRT we present in the experimental work section.

POSIX threads are used to allow the cores on each PC to compute concurrently while accessing shared-memory. Use of costly mutex-locks is kept to a minimum by processing the projection data in blocks that guarantee the threads perform voxel updates for non-overlapping image subvolumes. The only mutex-locks needed are for synchronizing the threads as they move from one set of projection blocks to the next. There are four such sets per projection.

OpenMPI based global reductions synchronize the computation across the PCs and tie together the distributed-memory defined thereby. The projection data blocks are distributed across the nodes in a round-robin fashion. Meanwhile, each node maintains a full copy of the image and other auxiliary

data structures. An image-sized global reduction is performed at the end of each iteration to complete image update vector:

$$\mathrm{u}^{(k)} = \mathrm{A}^{\mathrm{T}}\mathrm{R}(\mathrm{b} - \mathrm{A}\mathrm{x}^{(k)}). \tag{10}$$

During the first iteration, a second image-sized global reduction establishes preconditioning scalar $\mathrm{p} = 1/\|\mathrm{A}\|_1$. Following these global reductions, each local image copy is updated:

$$\mathrm{x}^{(k+1)} = (1 - \beta\,\mathrm{p})\,\mathrm{x}^{(k)} + \alpha\;\mathrm{p}\;\mathrm{MPI}(\Sigma\,\mathrm{u}^{(k)}). \tag{11}$$

Ordered subsets are handled by means of an outer loop that controls the extent of the forward and backprojections.

The system matrix is based on trilinear interpolation in the image space. The immense size of it necessitates that we compute it on-the-fly on a ray-by-ray basis. Repeatedly having to do so is costly. This part of the code is therefore vectorized to take advantage of Intel's multi-variable SSE instructions. The Siddon-like code for advancing along the projection ray is vectorized to simultaneously do so in the x, y, and z directions. The code for computing the trilinear interpolation coefficients at each step is vectorized in a more traditional sense, namely, by making four otherwise sequential arithmetic computations take place in parallel.

IV. HFIR CG-1D Cold Guide Beamline

HFIR produces high energy neutrons. After having been moderated, the neutrons travel down a long, slightly curved guide before they exit through a collimator which has an aperture of diameter D. The neutrons then travel a distance L as a focused beam till they reach the object which is placed on a rotating stage. Transmitted neutrons continue a short distance d before they hit an LiF/ZnS scintilator. The visible light produced thereby is bounced off a mirror and then recorded by a CCD camera. We use the term detector as a collective noun for the combined scintilator, mirror and CCD camera subsystem when the individual components need not be identified.

The effective resolution of the system depends on several factors. For example, a large L/D ratio leads to higher resolution as does a small value of d. See [1] for details. All three parameters have practical limits associated with them. The experimental work reported below was done using the settings $L = 5800$mm, $D = 8$mm, and $d = 100$mm. The values for distances L and d are hard to change due to physical constraints imposed by the building and the rotation stage, respectively. Aperture diameter D was chosen as a compromise between focusing the beam tightly and ensuring enough flux be available for the exposure time allocated to each projection. The latter was set by dividing the total beamtime available by the number of projections needed.

Data is typically acquired for several minutes for each view of the object. The longer the exposure time, the greater the likelihood that the neutron data is corrupted by gamma-ray strikes generated by the reactor core or by interactions between neutrons and elements in the CG-1D beamline itself. These random events are suppressed using an adaptive median filter which only replaces extreme values within its window. Other corrections needed include dark current subtraction, scaling of the data to compensate for variations of the beam intensity, and geometric alignment with the optical axis of the system.

V. Diesel Particulate Filter Imaging

Emissions regulations in the United States require that every on-road diesel vehicle be equipped with a filter that captures soot and other particulate. The filter can be described as a cylindrical object equipped with numerous parallel channels half of which are plugged in one end and the other half plugged at the other end. Made from porous material such as cordierite or silicon carbide, exhaust gases can flow through the filter walls and thus continue through the exhaust system but particulate matter gets trapped. Every so often, the filter must be run at a high temperature to regenerate (oxidize) the soot deposits. The logistical process of trapping and regenerating the soot is well-understood but there is no general understanding of how the deposits are distributed in the filter nor is it known how much fuel is required to adequately regenerate a filter. Adding to the complexity of the problem, a number of un-regenerable particulates accumulate thereby reducing both fuel efficiency and the effective volume of the filter. The interaction between soot and this "ash" is not well-understood either.

Destructive testing of a filter prevents longitudinal studies. Additionally, it is likely that the mechanical action of the cutting process will disturb the position and density of the particulate layers. Non-destructive testing is therefore desirable. One possibility is to use x-ray imaging. However, since both the filter and the particulate matter is hygroscopic (able to hold water), we have chosen to instead investigate use of neutron imaging. As an added benefit, this allows the filter to remain in its metal casing. A collection of filters have been loaded with varying amounts of soot and other particulate. Tomographic images are generated and analyzed over the course of several regeneration cycles. The analysis consists of segmentation and quantification of deposited matter.

VI. Experimental Results

We acquired 366 projections of an assembly holding two diesel particulate filters. Each projection was exposed to the neutron beam for 220 seconds. The angular increment between projections was 0.5 degrees for a total angular span of 180 degrees plus fan angle which satisfies the basic requirements for shortscan imaging. See Fig. 2 for an example of a projection after corrections and log-normalization. With transaxial detail being more important, all projections were axially downsampled a factor x16 to a size of 2048×128 resulting in 38×608 μm detector pixels.

PSIRT was executed for 36 iterations using 20 ordered subsets. We used $\mathrm{x}^{(0)} = 0$. Reconstruction was restricted to a cylinder with a radius of 0.85 of the support cone as it contained all the data. This effectively reduced the projections to a size of 1740×128 and resulted in an image of size $1642\times1642\times98$ with $40\times40\times800$ μm voxels. Trace analysis of the corresponding system matrix revealed a value of $\lambda^*_{\min} = 0.0005$ emphasizing the need for regularization. The largest column sum was 25.00 for the system matrix as a whole and

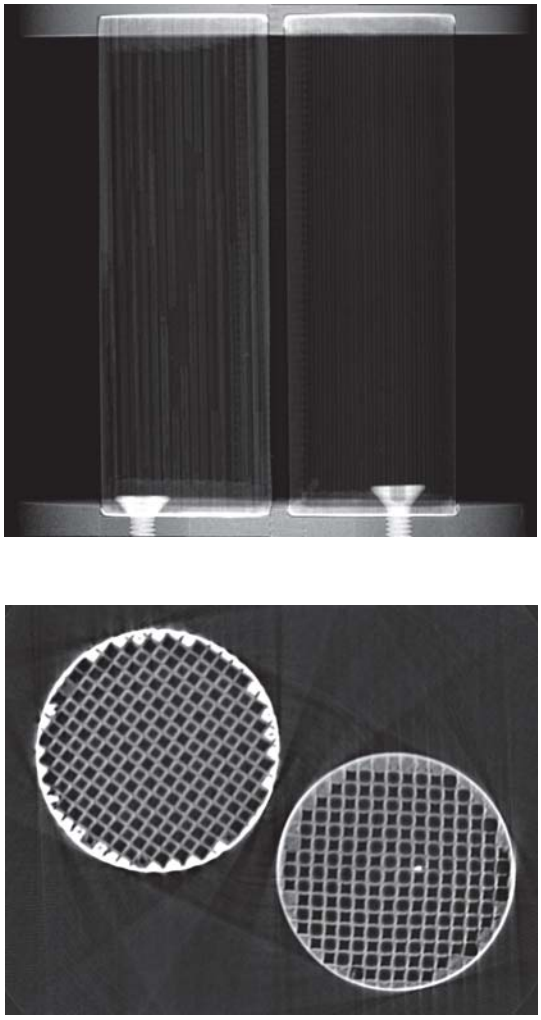


Fig. 2. Diesel particulate filter projection (top) and reconstruction (bottom).

1.25 for any one ordered subset. Accordingly, we used $\alpha = 1.95$ for relaxation and $\beta = 0.0125$ for regularization.

Figure 2 illustrates the quality of a reconstructed image which was cropped and contrast enhanced for illustrational purposes. We make the following observations. The internal filter structure is clearly visible. All channels appear square in the left filter but many are rounded to point of being almost circular in the right filter. This is not a reconstruction artifact but rather a reflection of the true geometry of the filters. The outer layer of the left filter is substantially brighter than for the one to the right. This is also not a reconstruction artifact but a result of material differences of the filters. The bright spot in the middle of the right filter is a form of particulate but it is not soot. The neutron cross-section for soot is very similar to that of the ceramic filter walls it attaches to making build-up difficult to discern visually. However, the reader may notice that some of the channels in the left image appear smaller than the channels adjacent to them. This subtle difference indicates the presence of particulate on those walls. We quantify build-up by segmenting the images and comparing the results over time as the filters go through the process of regeneration.

TABLE I
TIMING RESULTS FOR NEUTRON DPF APPLICATION

Iteration	Matrix	FBProj	MPI	Update	Total
k=0	78.7 s	353.8 s	85.8 s	4.4 s	522.7 s
k=1	78.7 s	256.6 s	43.0 s	4.4 s	382.7 s

To provide context for the image resolution and quality achieved, we point out that the linear system being solved is heavily under-determined with 81.5 million equations available to determine 198.8 million unknowns. Ideally, the situation should be the exact opposite, but that would have required taking more projections than possible within the beamtime allotted.

Table I summarizes timing results for the first (k=0) and second iteration (k=1) with the latter being representative of all subsequent iterations. The computer platform consisted of four Dell PCs equipped with dual quad-core Xeon processors running at 2.26 GHz. The network was 10G InfiniBand. We note that the elements of the system matrix ("Matrix") are computed in about a third of the time it takes to perform the forward and backprojections ("FBProj"). A single MPI based global reduction ("MPI") costs half as much again and does consequently not contribute much to the overall cost. The image update which includes relaxation and regularization takes negligible time ("Update"). The difference in cost between the first and all subsequent iterations is due to preconditioning scalar p being computed once, then stored and reused later. A total of 1.039 trillion interpolation coefficients are computed and used each full iteration. For the second iteration and onward, that translates into a computational cost of 23 clock cycles per coefficient which is very low. Moreover, it indicates that the code is cache and pipeline friendly and that further speedup most likely will have to come from an algorithmic reduction in the amount of data considered and/or use of more/faster PCs rather than from implementation improvements.

ACKNOWLEDGMENTS

Funding was provided by the U.S. Department of Energy, Office of Energy Efficiency and Renewal Energy, Vehicle Technologies Program under contract DE-AC05-00OR22725. Research at HFIR was partially sponsored by the Scientific Users Facility Division, Office of Basic Energy Sciences, U.S. Department of Energy.

REFERENCES

[1] K. Tobin, P. Bingham and J. Gregor, "Mathematics of Neutron Imaging," Chapter 7 in *Neutron Imaging and Applications*, Springer, 2009.
[2] A. Kak and M. Slaney, *Principles of CT Imaging*, SIAM, 2001.
[3] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Medical Imaging*, 27: 918–924, 2008.
[4] J. Gregor, "Distributed CPU multi-core implementation of SIRT with vectorized matrix kernel for micro-CT," *Fully 3D*, Germany, 2011.
[5] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1985.
[6] W. Press et al, *Numerical Recipes*, Cambridge Univ. Press, 2007.

Image Reconstruction for Proton CT Using Different Outputs

C. Bopp, M. Rousseau, D. Brasse
Université de Stasbourg, IPHC, 23 rue du Loess 67037 Strasbourg, France
CNRS UMR7178, 67037 Strasbourg, France

***Abstract*—Proton CT (pCT) nowadays aims at improving hadron therapy treatment planning by mapping the electron density of materials. The main information used is the energy of the protons. However, during a pCT acquisition, the spatial and angular deviation of each particle is recorded and the information about its transmission or not is implicitly available. The pCT scan of a realistic head phantom with tumours of different chemical compositions but the same electron density was simulated and images were reconstructed using these observables. The first reconstructions do not allow to conclude on the potential determination of chemical composition using these observables yet. Nevertheless, preliminary results indicate these observables could bring information of interest.**

I. Introduction

In the last two decades, a great interest has been shown for hadrontherapy, as many centres for proton or 12-C therapy have opened. These particles offer the advantage of allowing a more localized dose deposition than conventional radiotherapy, delivering a higher dose to the lesion, while sparing the tissues around. However, this interest has resulted in the need for more precise tools for treatment planning. Treatment plans established nowadays are based on X-ray Computed Tomography images and are of limited accuracy because of the difference in the nature of the interactions between photon and protons or carbon ions. The acquired CT numbers are converted into water-equivalent stopping powers using a calibration curve [1]. It has been shown that the uncertainty on CT numbers due to beam hardening effects as well as inaccuracies due to the calibration curve used for the conversion can induce a shift in the depth of the Bragg peak up to 2% [2]. This translates to a range uncertainty between 1 and 3 mm in typical treatment situations.

Proton imaging has been put forward as a way to address this problem: the principle of pCT is the mapping of the relative electron density of the materials with respect to water, and thus the stopping power of the tissues protons went through, by using the information on their energy loss. Protons of energy high enough to have the Bragg peak after the object to image are generated. The position and direction of each particle before and after the object are recorded, as well as the exit energy of the protons [3]. The aim is to be able to map the relative electron density of the materials with respect to water, and thus the stopping power of the tissues protons passed through, by using the information on their energy loss, through the Bethe-Bloch theory. Therefore one of the main parts of a proton scanner is the calorimeter. However, the main challenge was long considered to be the poor spatial resolution issue due to MCS. The use of the trackers, coupled to a list-mode acquisition, allows the determination of the path of individual protons for image reconstruction. While the fastest solution as far as image reconstruction is concerned is the straight line approximation, it results in a decrease in the spatial resolution. The use of the Most Likely Path [4], computed using the entry and exit position and direction of each particle, supposing the object is made of water, currently gives the most accurate results [5].

Over the years, a few studies have taken different approaches about proton imaging. For example, it was proposed to use the nuclear scattering of 500 to 1000 MeV protons to obtain a three-dimensional image reconstruction of an object in one exposure and getting information on densities as well as hydrogen concentrations of materials [6]. Proton radiography using MCS has also being evoked [7], although for its edge-delineation properties.

In a similar fashion, other studies have put forward different potential sources of inaccuracies or important effects to consider during hadrontherapy treatment planning other than the range uncertainty. The influence of the chemical composition of the materials on the dose deposit, and the interest of assigning nuclear interaction cross sections to the materials for treatment planning was studied [8]. It was shown that the conversion of CT number without taking into account the composition of the materials could induce a difference in the maximal energy deposit up to 1.5% for a 12-C beam in materials presenting the same CT numbers [9].

The aim of this study is to explore the possibility to acquire different information about the materials using the data available from a pCT scan. The possible differentiation of materials, some presenting the same electron density, will be investigated.

II. Materials and Methods

A. Outputs of a pCT scan and observables

During a pCT scan, the data recorded for each particle are the following:

- the energy after the object. Supposing the beam energy is known, this gives us access to the energy loss used to calculate the water equivalent path length of each particle. This makes it possible to reconstruct an image of the electron density of the materials.

- the positions and directions before and after the object to image.

Through these data, we have defined observables of the interactions of protons with the matter at our disposition after a pCT scan. The information on the energy also gives access to the information on the straggling of the energy loss, which is described by Bohr's [10] and Tschalar's [11], [12] theories. However, the variations of this observable with different materials are very small compared to the potential resolution of a calorimeter. Because of that, it was not studied here.

The information from the trackers gives us access to the spatial and angular deviation of each particle, due to MCS. This information is used in order to calculate the most likely path (MLP) of each particle and improve the spatial resolution during image reconstruction. The distributions of the spatial and angular deviation are proportional, and depend mainly on the energy of the particles and the radiation length of the materials. Because of this, it could help differentiate materials. Because of its invariance with respect to the position of the trackers, the angular deviation was the observable we kept for this study.

As the trackers record the positions and directions of each particle before and after the object, the information on the transmission or not of each particle is also implicitly available. The attenuation of the particles is due to non-elastic nuclear interactions, and the reconstruction of transmission images could bring information on the interaction cross section of protons in the materials.

A previous study has showed that the information on the angular deviation and transmission rate could bring complementary information about the materials, and might be used to detect hypoxia [13]. The statistical uncertainty on these two observables is higher than the one on the energy, indicating the pertinence to study a voxel or a region of interest (ROI) in a reconstructed image. This makes it possible to get access to high statistics while limiting the dose delivered to the potential patient.

B. Simulation of a pCT scan of a realistic head phantom

Simulated pCT scan data were generated using the Gate platform [14]. The phantom used was a dedicated MRI head phantom [15]. To each part of the head, a material was attributed. Two lesions were also inserted inside the brain: two carcinoma with different ratios of oxygen but the same electron density. Mono-energetic 200 MeV protons were sent from a plane source, all along the same direction (Figure 1). A set of 256 projections was simulated with 100 protons per square millimetre. Secondary protons created from nuclear interactions were not considered in the analysis. The dose delivered for the acquisition amounts to 2.5 mGy.

C. Pre-processing of the observables

1) From energy to electron density: At the energies considered of pCT (typically 200 MeV for a head scan, 250-300 MeV for a torso scan) the energy loss of the protons through inelastic collisions with electrons is described by the Bethe-Bloch theory and can be written as follows in order to reconstruct an image:

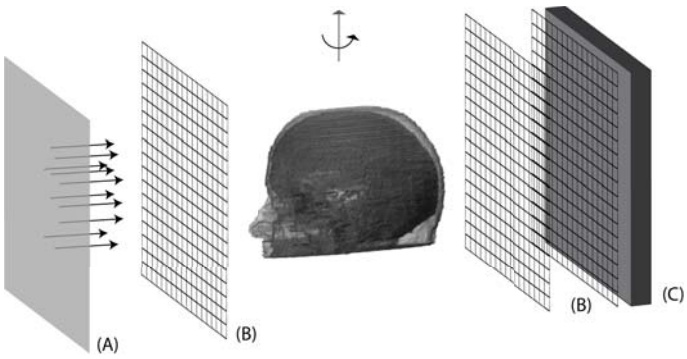


Fig. 1: pCT scanner system simulated. (A) Parallel proton source, (B) Trackers, (C) Calorimeter.

$$\int_{E_{in}}^{E_{out}} \frac{dE}{F(I_{water}, E)} = \int_L \eta_e(r) dl \tag{1}$$

where E_{in} and E_{out} are the energies of the proton before and after the object, and F a function depending on the proton energy and the water ionization potential I_{water}. L is the proton path in the object and η_e the relative electron density (or stopping power) with respect to water.

This data was binned into projections according to the entry positions of the protons as well as from the median position of the particles on the MLP.

2) Pre-processing for transmission imaging: Non-elastic nuclear interactions lead to a reduction of proton fluence with increasing thickness of the object. The attenuated fluence of protons can be expressed as follows in the continuous slowing down approximation:

$$\Phi(r) = \Phi_0 \exp\left(-\int_L \kappa(r) dl\right) \tag{2}$$

in the materials. This lead us to reconstruct a transmission image, in a similar way to X-ray CT.

The list-mode data was binned into projections according to the entry position of the particles with in each pixel, the value V:

$$V = -\log \frac{N}{N_0} \tag{3}$$

with N and N_0 respectively the number of protons sent and detected.

3) Pre-processing for the deviation – integration along the most likely path: The deviation angle of each particle is only representative of the materials encountered in a statistical way. In order to keep the advantage of each proton carrying its own information, we quantified the deviation of each proton. As particles can exit with similar deviation angles but different positions (Fig. 2), the information on the deviation was not used directly. We have calculated the most likely path for each proton, and integrated the deviation along the path as shown on Fig. 3.

This data was then binned into projections according to the source position of the particles.

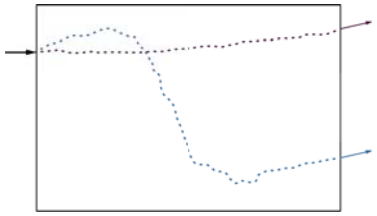


Fig. 2: Two protons with the same deviation angle but different histories and exit positions

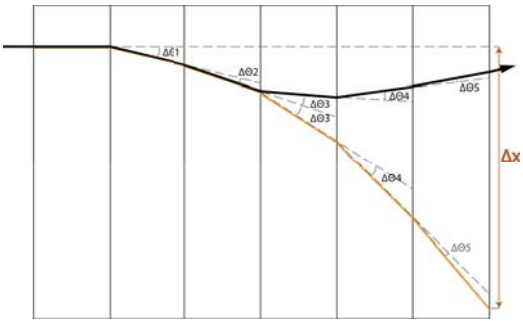


Fig. 3: Integration of the deviation along the most likely path

D. Reconstruction algorithms

Different algorithms were tested to reconstruct the pre-processed data binned into projections: the Filtered Back-projection (FBP) algorithm, the Algebraic Reconstruction Technique (ART) and the Maximum-Likelihood Expectation-Maximization (ML-EM). The ART + Total Variation (TV) was also used, both on projections and on the list-mode data for the reconstruction of the relative electron density.

E. Contrast study of the images

Three ROIs were defined in the reconstructed images as follows: one inside each carcinoma and one in the brain tissue, as shown on Figure 4. The contrast in these regions was studied using the following Figure of Merit (FOM):

$$\mathrm{FOM} = \frac{\overline{\mathrm{BrainROI}} - \overline{\mathrm{LesionROI}}}{\overline{\mathrm{BrainROI}}} \tag{4}$$

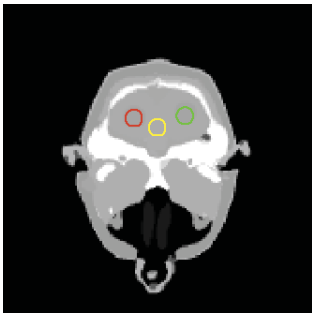
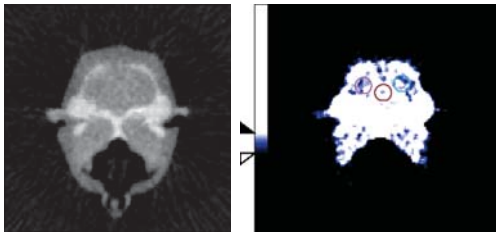


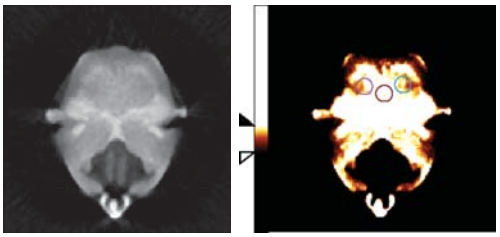
Fig. 4: Transverse slice of the phantom. In red (top left) and green (top right) ROIs inside the carcinoma tissues, in yellow (bottom) a ROI taken in the brain tissue.

(a) Relative electron density radiography (b) Log transmission radiography (c) Angular deviation radiography

Fig. 5: Radiographies of the head phantom using different observables



(a) Transmission



(b) Integrated deviation

Fig. 6: Transmission and Integrated deviation images reconstructed with FBP

III. Preliminary Results

A. Projections using different observables

The projections at one acquisition angle for the three pre-processed data (relative electron density, transmission and deviation) are shown in Fig. 5. The quantity of information contained in the transmission and deviation is important. The low statistics used to generate the projections (100 protons/mm^2) results in a low Signal to Noise Ratio (SNR), for the transmission and for the integrated deviation.

B. Reconstructions

The ART algorithm alone does not show good results with noisy data, and thus was not appropriate for the reconstruction of the transmission and deviation. FBP reconstructions of the transmission and deviation are shown in Fig. 6. The SNR ratio in both images is quite low. Qualitatively, the different soft tissues are hardly differentiable. Nevertheless, the differences due to the lesions in the brain can be made apparent by changing the contrasts in the image.

C. Contrast study

The values of the contrast FOM for the two carcinoma are shown in Table I for the different observables and reconstruction algorithms. The contrast is higher for both the transmission and the deviation image than for the electron density.

TABLE I: Contrast FOM as defined in equation (4) between the two carcinoma and the brain ROIs. Results shown for images binned according to the entry position of the protons.

		Left carcinoma	Right carcinoma
Electron density	FBP	3.1%	2.3%
	ML-EM	2.8%	2.1%
	ART+TV	2.9%	2.3%
Transmission	FBP	11.2%	9.8%
	ML-EM	14.1%	12.6%
Integrated deviation	FBP	9.7%	12.0%
	ML-EM	10.9%	12.2%

IV. Discussion and Conclusion

The information on the transmission and deviation of the particles acquired during a pCT scan is not yet used to its full potential. The head radiographies using the different outputs shown on Fig. 5 indicate that the quantity of information carried by the transmission and deviation is important. The SNR of the radiographic projections suggest that reconstructions will be noisy. The FBP reconstructions shown on Fig. 6 confirm this. Nevertheless, the two carcinoma can be differentiated from the brain using both the transmission and the deviation. The contrast study shown in Table I indicates a higher contrast between the lesions and the brain in the images of transmission and deviation than in the images of the electron density. This could be of interest in diagnostics. However, the low SNR in the images does not allow to conclude yet on the potential detection of hypoxia. The differences in contrast between the left and right carcinoma are too small compared to the noise in the images. In order to get more information on the materials, more appropriate reconstruction algorithms need to be developed. As an example, while binning the data according to the entry position of the particles seems natural as far as the transmission information is concerned. It has been shown that it is not the most efficient as far as the electron density is concerned. Results indicate that a more precise approximation of the particles paths should be considered when reconstructing the deviation of the protons. Indeed, one can notice that there is a slight discrepancy between the positions of the lesions and the localization of the effect of this difference in the reconstruction of the deviation (Fig. 6b). A reconstruction using the list-mode data and the MLP for each proton should give better results.

On-going studies include the quantification of the effect of the binning (entry position vs. median position on the MLP), the effect of the statistics, as well as the effect of cuts on the exit angles in order to eliminate particles that underwent nuclear scattering on the reconstruction of the deviation. These cuts are expected to greatly improve the SNR in the deviation images.

This study is a first step towards the possibility to identify the chemical composition of materials using proton imaging. One possible research axis for further work could be the development of a reconstruction algorithm taking into account more than one of the observables.

Acknowledgements

This work was partly supported by the French National Research Agency through the ProTom project (ITMO Cancer et Technologie pour la santé - pc201123).

References

[1] U. Schneider, E. Pedroni, and A. Lomax, "The calibration of ct hounsfield units for radiotherapy treatment planning," *Physics in Medicine and Biology*, vol. 41, no. 1, p. 111, 1996.

[2] B. Schaffner and E. Pedroni, "The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between ct-hu and proton stopping power." *Phys Med Biol*, vol. 43, no. 6, pp. 1579–92, 1998.

[3] R. Schulte, V. Bashkirov, T. Li, Z. Liang, K. Mueller, J. Heimann, L. R. Johnson, B. Keeney, H. F. W. Sadrozinski, A. Seiden, D. C. Williams, L. Zhang, Z. Li, S. Peggs, T. Satogata, and C. Woody, "Conceptual design of a proton computed tomography system for applications in proton radiation therapy," 2004.

[4] D. C. Williams, "The most likely path of an energetic charged particle through a uniform medium," *Phys. Med. Biol.*, vol. 49, no. 13, p. 2899, 2004.

[5] T. Li, Z. Liang, J. Singanallur, T. Satogata, D. Williams, and R. Schulte, "Reconstruction for proton computed tomography by tracing proton trajectories: a monte carlo study." *Med Phys*, vol. 33, no. 3, pp. 699–706, 2006.

[6] J. Saudinos, G. Charpak, F. Sauli, D. Townsend, and J. Vinciarelli, "Nuclear scattering applied to radiography," *Physics in Medicine and Biology*, vol. 20, no. 6, p. 890, 1975.

[7] J. Cookson, "Radiography with protons," *Naturwissenschaften*, vol. 61, 1974.

[8] H. Palmans and F. Verhaegen, "Assigning nonelastic nuclear interaction cross sections to hounsfield units for monte carlo treatment planning of proton beams," *Physics in Medicine and Biology*, vol. 50, no. 5, p. 991, 2005.

[9] E. Batin, "Influence de la composition chimique des tissus humains sur les dpts de dose en hadronthrapie," Ph.D. dissertation, Universit de Caen, 2008.

[10] N. Bohr, *The penetration of atomic particles through matter*. Munksgaard, 1948.

[11] C. Tschalar, "Straggling distributions of large energy losses," *Nuclear Instruments and Methods*, vol. 61, no. 2, pp. 141 – 156, 1968.

[12] ——, "Straggling distributions of extremely large energy losses," *Nuclear Instruments and Methods*, vol. 64, no. 3, pp. 237 – 243, 1968.

[13] C. Bopp, J. Colin, C. Finck, M. Labalme, M. Rousseau, and D. Brasse, "Observable analysis for proton computed tomography," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, oct. 2012.

[14] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt *et al.*, "Gate v6: a major enhancement of the gate simulation platform enabling modelling of ct and radiotherapy," *Physics in Medicine and Biology*, vol. 56, no. 4, p. 881, 2011.

[15] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. R. Gindi, and P. B. Hoffer, "Computerized three-dimensional segmented human anatomy," *Medical Physics*, 1994.

Conference Appearance Index

Martin Fiebich WeP2
Michael Fieseler MoO3
Andreas Fieselmann TuO2
Charles Finney ThP3
Barbara Flach TuO1
Gilles Fleury MoO1
Sascha Fränkel ThP3
Erik Fredenberg WeO3
Jurgen Frikel ThP3
Geng Fu ThP3
Lin Fu TuO1, TuO1
Theobald Fuchs WeO3

G

Hao Gao MoP1, WeP2
Pascal Getreuer ThP3
Pieter Ghysels WeP2
Jens Gibmeier ThP3
Ziya L. Gokaslan TuO1
Mahsa Golkar MoP1
Marlies C. Goorden MoP1
Robert G. Gould TuO2
Yves Goussardy MoP1
Michael Grass MoP1
Jens Gregor ThP3
Daniel Groß ThP3
Grant T. Gullberg MoP1, TuO2, WeP2, ThO1, ThP3
Donald L. Gunter WeP2
Rajiv Gupta ThP3

H

Sungsoo Ha MoP1
Abdelmoula Haboub WeO3
Dieter Hahn ThO2
Hao Han ThP3
Xiao Han MoP1, WeP2, WeP2
Jia Hao MoP1
Hein Heidbüchel MoP1
Ulrich Heil ThP3
Jürgen Hesser MoP1
Joachim Hornegger MoO2, MoP1, MoP1
WeO2, ThO3, ThP3
Jiang Hsieh WeP2, ThP3
Scott Hsieh ThP3
Qiu Huang MoP1
Brian F. Hutton WeP2, WeP2

J

David Jaffray ThP3
Abhinav K. Jha MoP1
Yannan Jin ThP3
Karen L. Johnson WeP2, ThO1
Roger H. Johnson WeP2
Jakob S. Jørgensen MoP1, TuO1

K

Marc Kachelrieß MoO1, MoO2, TuO1, WeO1, WeP2
Willi A. Kalender ThP3
Kalpana M. Kanal WeO3
Kejun Kang MoP1
W. Clem Karl ThO3, ThP3
Joel S. Karp ThO1
Marc Käseberg MoP1, ThP3
Alexander Katsevich WeP2
Daniil Kazantsev WeP2
Benjamin Keck WeO1
Erwin Keeve MoP1, ThP3
A. Jay Khanna TuO1
Donghwan Kim MoO1
Paul E. Kinahan WeO3, WeP2, ThO1
Michael A. King WeP2, ThO1, ThP3
Keishi Kitamura ThO3
Gerhard Kleinszig TuO1
Michael Knaup MoO1, WeP2
Michal Knas ThP3
Tetsuya Kobayashi ThO3
Thomas Koehler ThO3
Christoph Köhler MoP1
Arda Könik ThO1, ThP3
Thomas Kosters MoO3
Markus Kowarschik TuO2
Hiroyuki Kudo ThO3
Rolf Kueres TuO1
Jan Kuntz TuO1
Holger Kunze ThP3
Matthew A. Kupinski MoP1
Meredith Kupinski ThP3
Yiannis Kyriakou MoP1, ThP3

L

Patrick J. La Rivière TuO2
Alexander Ladikos MoO2
Elisabeth Lahalle MoO1
Helene Langet MoO1
Tobias Lasser ThP3, ThP3
Günter Lauritsch MoO2, MoP1, WeO2, ThP3
Richard M. Leahy TuO2, WeP2
Taek-Soo Lee MoP1
Taewon Lee WeP2
Yin-Jie Lei MoP1
Hao Li ThO1
Hongwei Li MoP1
Lei Li WeP2
Liang Li ThP3
Quanzheng Li TuO2, WeP2, ThO1
Yusheng Li ThO1
Zhengrong Liang MoP1, ThP3
William R. B. Lionheart WeP2
Kevin J. Little TuO2
Baodong Liu WeP2
Feng Liu MoP1
Fenglin Liu MoP1
Xuan Liu ThP3
Yan Liu MoP1, ThP3
Vladimir Lobastov ThP3
Yong Long ThO1
Daniel A. Low WeP2

Hongbing Lu MoP1, ThP3

M

Jianhua Ma ThP3
Nicole Maass WeO_3, ThP3
Lawrence R. MacDonald WeO_3
Alastair A. MacDowell WeO_3
Andreas K. Maier MoP1, MoP1, TuO_2, WeP_2, ThP3
Jens Maisenbacher ThP3
Michael Maisl MoP1
Andreas Malecki ThP3
Jonathan S. Maltz WeP_2
Joseph J. Manak WeP_2
Michael T. Manhart TuO_2, WeP_2
Stefano Marchesini WeO_3
Harry E. Martz Jr. MoP1
Samuel Matej ThO_1
Madison G. McGaffin TuO_1, WeP_2
Kai Mei WeP_2
Deirdre R. Meldrum WeP_2
Steffen Melnik ThP3
Scott D. Metzler ThO_1
Chuang Miao ThP3
Lucian Mihailescu WeP_2
Gregory S. Mitchell MoP1
Debasis Mitra ThO_1
Dimple Modgil TuO_2
Xuanqin Mou MoP1, ThO_2, ThO_3
Klaus Mueller MoP1, WeO_1
Kerstin Müller MoO_2, MoP1, ThP3

N

Hye Sun Na WeP_2
Radin A. Nasirudin WeP_2
Brian E. Nett WeO_2, WeP_2, WeP_2, ThP3
Hung Nien WeP_2
Peter B. Noël WeP_2, ThO_2
Frédéric Noo WeO_2, ThO_3
Dieter Nuyens MoP1
Johan Nuyts MoO_3, WeP_2

O

Michael O'Connor WeP_2
Dan Oberg WeP_2
Steven Oeckl MoP1
Markus Oehlhafen MoO_2
Yoshito Otake TuO_1
Sebastien Ourselin WeP_2, WeP_2

P

Jed D. Pack WeO_2, WeP_2, WeP_2, ThP3
Atul Padole ThO_3
Debashish Pal TuO_1
Victor P. Palamodov ThP3
Xiaochuan Pan MoP1, MoP1, MoP1, TuO_1
WeP_2, WeP_2, ThP3
Vladimir Y. Panin MoO_3
Eric Papenhausen WeO_1
Nikos Paragios MoO_1
Diworth Y. Parkinson WeO_3
Javier Pavía ThP3
Pascal Paysan MoO_2
Erik A. Pearson MoP1, WeP_2
Stefano Pedemonte WeP_2
Charles A. Pelizzari MoP1, WeP_2
Petar Penchev WeP_2
Franz Pfeiffer ThO_2, ThP_3, ThP3
Larry A. Pierce II WeO_3
Francisco Pino ThP3
Guillaume Potdevin ThP3
P. Hendrik Pretorius WeP_2, ThO_1, ThP3
Yi-Fei Pu MoP1

Q

Jinyi Qi MoO_3, MoP1
X. Sharon Qi WeP_2
Gangrong Qu WeP_2

R

Sathish Ramani MoO_1
Ran Ren WeP_2
Aymeric Reshef MoO_1
Christian Reuß WeO_3
Ahmadreza Rezaei MoO_3, WeP_2
Cyril Riddell MoO_1
Christian Riess WeP_2
David Rigie TuO_2
Ludwig Ritschl MoO_1
Christopher Rohkohl MoO_2, MoP1
Domenec Ros ThP3
Jeffrey M. Rosen WeO_1
Marc Rousseau ThP3
Xue Rui ThO_1
Ernst J. Rummeny WeP_2

S

Adrian A. Sanchez ThP3
Alexander Sasov ThP3
Ken D. Sauer TuO_1, ThO_2, ThP3
Stefan Sawall WeP_2
Dirk Schäfer MoP1
Klaus Schafers MoO_3
Ingo Schasiepen ThP3
Katharina Schmitt ThO_3
Stephen M. Schmitt MoP1
Elmar Schömer ThP3
Tobias Schon WeO_3
Harold Schöndube ThO_3
Christian Schorr MoP1
Ralf Schulze ThP3
Ulrich Schwanecke ThP3
Chris Schwemmer MoO_2, MoP1
W. Paul Segars ThO_1
Youngho Seo TuO_2
Mohammed S. Shazeeb ThO_1
Le Shen WeP_2
Daxin Shi MoP1
Uttam Shrestha TuO_2
Emil Y. Sidky MoP1, MoP1, TuO_1, WeP_2, ThP3

Jeffrey H. Siewerdsen TuO1, WeO2, WeO3, ThO3
Jan Sijbers WeP2
Sarabjeet Singh ThO3
J. Webster Stayman TuO1, WeO2, WeO3, ThO3
Enrico Stefani ThP3
Dzmitry Stepankou MoP1
Karl Stierstorfer ThO2, ThO3
Fabian Stopp MoP1
Olivier Strauss WeO1
Johan Sunnegardh ThO2
Marc Sussman ThO3

T

Katsuyuki Taguchi MoO2, WeO3
Thomas M. Talavage MoO1
Qiulin Tang MoO2
Shaojie Tang ThP3
Xiangyang Tang ThP3
Hideaki Tashima WeP2
Arthur Tenenhaus MoO1
Jean-Baptiste Thibault TuO1, TuO1, ThO2, ThP3
Pierre Thibault ThO2
William M. Thompson WeP2
Mengqiu Tian ThP3
Todd Toops ThP3
Yves Trousset MoO1
Benjamin M. W. Tsui MoO1, MoP1

V

Wim van Aarle WeP2
Peter van de Haar MoP1
Frans van der Have MoP1
Michael W. Vannier MoP1
Wim Vanroose WeP2
Phillip A. Vargas TuO2
Alexander I. Veress ThP3
Kai Vetter WeP2
Dimitris Visvikis WeO1, WeP2
Jakob Vogel ThP3
Sebastian Vogt TuO1
Matthieu Voorons MoP1

W

Adrien Wagner MoP1
Katherine L. Walker MoP1
Adam S. Wang TuO1, ThO3
Ge Wang MoP1, ThO3, ThP3
Guobao Wang MoO3
Jing Wang ThP3
Ke Wang MoP1
Linyuan Wang WeP2
Qingli Wang ThP3
Alexander Wanner ThP3
Wolfgang Wein MoO2
Oliver Weinheimer ThP3
Thomas F. Wenisch WeO1
Julia Wicklein ThP3
Adam J. Wieckowski MoP1
Matthias Wieczorek ThP3
Haibo Wu WeP2
Junfeng Wu ThO3
Junjie Wu WeO1
Meng Wu WeP2
Xundong Wu ThP3
Yong Wu ThP3
Frank Wubbeling MoO3
Katrin Wunder ThP3

X

Xiaoqi Xi WeP2
Ming Xia ThP3
Yan Xia MoP1
Ya-Hong Xie ThP3
Yuxiang Xing WeP2
Jingyan Xu MoO1
Qiong Xu MoP1
Wei Xu MoP1

Y

Taiga Yamaya WeP2
Bin Yan WeP2
Zhongbo Yan ThP3
Kai Yang MoP1
Liu Yang ThP3
Meng-long Yang MoP1
Qiao Yang ThP3
Zhye Yin WeP2, WeP2
Hengyong Yu WeP2, ThO3, ThP3
Zhicong Yu WeO2
Zhou Yu TuO1, TuO1, ThP3

Z

Stanislav Zabic ThO3
Yunlong Zan MoP1
Wojciech Zbijewski WeO3
Chengeng Zeng WeO3
Kai Zeng WeO2, WeP2, WeP2
Hanming Zhang WeP2
Hao Zhang MoP1, ThP3
Huitao Zhang MoP1
Li Zhang MoP1, ThP3
Peng Zhang MoP1
Ruoqiao Zhang ThO2
Wei-Hua Zhang MoP1
Yanbo Zhang MoP1, ThO2
Yi Zhang MoP1
Zheng Zhang WeP2
Mengliu Zhao MoP1
Zhe Zhao ThO3
Ziyi Zheng MoP1, WeO1
Ji-Liu Zhou MoP1
Jian Zhou MoP1
Linghong Zhou ThP3
Wentao Zhu TuO2
Yining Zhu MoP1
Timo Zinsser WeO1